

AD\_\_\_\_\_

AWARD NUMBER: W81XWH-07-1-0492

TITLE: Interdisciplinary Studies on the Combat Readiness and Health Issues Faced by Military Personnel

PRINCIPAL INVESTIGATOR: John Hart, Jr., M.D.

CONTRACTING ORGANIZATION: University of Texas at Dallas  
Richardson, TX 75080

REPORT DATE: September 2010

TYPE OF REPORT: Final Addendum

PREPARED FOR: U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;  
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

<b>REPORT DOCUMENTATION PAGE</b>				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b>					
<b>1. REPORT DATE</b> 1 September 2010		<b>2. REPORT TYPE</b> Final Addendum		<b>3. DATES COVERED</b> 1 Sep 2009 – 31 Aug 2010	
<b>4. TITLE AND SUBTITLE</b>  Interdisciplinary Studies on the Combat Readiness and Health Issues Faced by Military Personnel				<b>5a. CONTRACT NUMBER</b>	
				<b>5b. GRANT NUMBER</b> W81XWH-07-1-0492	
				<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b>  John Hart, Jr., M.D.  E-Mail: jhart@utdallas.edu				<b>5d. PROJECT NUMBER</b>	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> University of Texas at Dallas Richardson, TX 75080				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>	
				<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b> Approved for Public Release; Distribution Unlimited					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT</b>  The goal of this research project was twofold:  Task # 1a: Assemble multimodal human performance laboratory including complex human motor assessment system, 128 channel EEG/ERP, pupillometer/eyetracking system, and repetitive transcranial magnetic stimulation system.  Task # 1b: Conduct a pilot research study demonstrating the capabilities of performing multimodal assessment of object retrieval, particularly when those objects may be considered threatening or nonthreatening.  Task # 1a has been accomplished and reported previously. The pilot study (Task #1b) is accomplished, results of which are reported in this document.					
<b>15. SUBJECT TERMS</b> rTMS, threat, PTSD, EEG, fMRI					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>  UU	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b> USAMRMC
<b>a. REPORT</b> U	<b>b. ABSTRACT</b> U	<b>c. THIS PAGE</b> U			<b>19b. TELEPHONE NUMBER</b> (include area code)

## Table of Contents

	<u>Page</u>
Introduction.....	4
Body.....	4
Key Research Accomplishments.....	5
Reportable Outcomes.....	5
Conclusion.....	6
References.....	6
Appendices.....	7

## **INTRODUCTION:**

The goal of this research project was twofold:

Task # 1a: Assemble multimodal human performance laboratory including complex human motor assessment system, 128 channel EEG/ERP, pupilometer/eyetracking system, and repetitive transcranial magnetic stimulation system.

Task # 1b: Conduct a pilot research study demonstrating the capabilities of performing multimodal assessment of object retrieval, particularly when those objects may be considered threatening or nonthreatening.

## **BODY:**

Task #1a. has been accomplished and reported previously. The Multi-Modal Brain-Motor Performance Laboratory is fully constructed and operational at the University of Texas at Dallas Center for BrainHealth located at 2200 W. Mockingbird Lane, Dallas, Texas 75235 in room 3.120. Additional planned components (a high-end immersive driving/task simulator and an eight camera human motion capture and analysis system) were also constructed and are located at the University of Texas at Arlington's Human Performance Institute, Nedderman Hall, Room 241.

We have recently also added to the Multi-Modal Brain-Motor Performance Laboratory two motion detection cameras that with sensors attached to a subject's knees and ankles can record human gait patterns for analysis.

The pilot study (Task #1B.) This has been conducted after a prolonged delay in IRB approval. Since receiving approval from the USAMRMC HRPO, we have conducted the pilot study in 5 normal controls while combining rTMS at 1 Hz for 20 minutes. Prior to and after the rTMS, the subjects were administered the visual object processing ERP/EEG measures that detect hyperarousal to threatening and nonthreatening pictures. These were performed successfully both prior to and after the rTMS procedures for every subject. There were no complications or difficulties experienced in performing the pilot research study

In addition, with equipment purchased from these grant funds, we have developed a study/technique where we have incorporated pupilometer measures of the same stimuli presented for the ERP/EEG measures. This provides for individual subject measures of their arousal responses for *individual stimuli*. This has now been performed on 12 volunteer controls and we are continuing to enroll individuals for this normalization study. This has also been addressed from the aspect of adjusting for the luminance of the stimuli (see attached technical report). By being able to perform assessments of each individual's arousal/hyperarousal responses, thus allowing for individualized assessment to specific hyperarousal stimuli. This advance better suits the system to perform individual assessments and detect treatment responses to specific arousing stimuli for a given individual.

The findings from the pilot studies of effectiveness of the rTMS-EEG system and the addition of pupilometry to perform individualized assessments of arousing stimuli has ideally positioned us to perform the studies proposed in our newly awarded CDMRP DMRDP grant entitled "Novel Treatment of Emotional Dysfunction in PTSD," that we have just received notification of award to the PI, Dr. John Hart, Jr. This new 3 year grant will use 12 sessions of rTMS at 1 Hz with Cognitive Processing Therapy administered immediately after each rTMS session. The hypothesis is that the rTMS will temporarily reduce the threat/hyperarousal response so the Cognitive Processing Therapy can have maximal effect. The present project provided the equipment and preliminary data that allowed for this treatment trial to be funded, for which we are exceedingly grateful.

## **KEY RESEARCH ACCOMPLISHMENTS:**

Completion of the integrated Multimodal Human Performance Laboratory, including:

- Repetitive Transcranial Magnetic Stimulation System with Magstim Super Rapid package and BrainSight Software System
- SensoMotoric Instruments (SMI) Eyetracker system
- Biologics EEG System
- Human Motor Performance assessment system

We have upgraded the system further by purchasing on our own an EMG unit to integrate with the other equipment.

We have also produced the following research developments:

1. Working on technological development in time-frequency EEG analysis which has lead to a recent publication (Ferree et al., Neuroimage, 2009).
2. Working with engineering on inspection of the rTMS unit and upgrade and integration with the EMG unit added to the multi-modal laboratory.
3. Established EEG markers of selection and inhibition of concepts and semantic memory. This lead to ERP markers (Maguire et al., 2009), and time-frequency and coherence measures (Brier et al., in press) of selecting correct and inhibiting incorrect memories.
4. We are in the process of validating a pupillometry measure with the established EEG measure of arousal/threat that allows for assessment of an individual's arousal response to specific emotional stimuli.

## **REPORTABLE OUTCOMES:**

Task #1a. has been performed.

Task #1b. has been performed.

This has culminated in the award of the CDMRP DMRDP grant entitled "Novel Treatment of Emotional Dysfunction in PTSD," to the PI, Dr. John Hart, Jr.

In addition, the following relevant presentations and publications in peer reviewed journals have

Presentations (J. Hart):

Ferree, T.C., Brier, M.R., Hart, J., Kraut, M.A. Space-Time-Frequency analysis of EEG data in single subject and group data. *Cognitive Neuroscience Society Meeting*, March 21 – 24, 2009.

Calley, C.S., Brier, M.R., Ferree, T.C., Patel, R.H., Gelman, Y., Tillman, G.D., Kraut, M.A., Hart, J. EEG correlates of processing visual threatening stimuli. *Cognitive Neuroscience Society Meeting*, March 21 – 24, 2009.

Brier, M.R., Maguire, M.J, Ferree, T.C., Gelman, Y., Patel, R., Hart, J., Kraut, M.A. Arithmetic Processing and Number Comparison as Revealed by Event Related Potentials. *Society for Neuroscience Meeting*, 2009.

Calley, C., Brier, M., Ferree, T., Patel, R., Gelman, Y., Tillman, G., Kraut, M., Hart, Jr., J. EEG correlates, of threatening visual stimuli. *Cognitive Neuroscience Society Annual Meeting*, Montréal, Canada, 2010.

Kraut, M., Brier, M., Maguire, M., Ferree, T., Hart, J. Frontal theta and alpha power and coherence changes are modulated by semantic complexity in Go/NoGo tasks. *Cognitive Neuroscience Society Annual Meeting*, Montréal, Canada, 2010.

Papers (J. Hart):

Ferree, T., Brier, M., Hart, J., Kraut, M. Space-time frequency analysis of EEG data using within-subject statistical tests followed by sequential PCA. *Neuroimage*, 45(1):109-21, 2009.

Maguire, M., Brier, M., Moore, P., Ferree, T., Ray, D., Mostofsky, S., Hart, J., Kraut, M. The influence of perceptual and semantic categorization on inhibitory processing as measured by the N2-P3 response. *Brain and Cognition*, 71, 59, 2009.

Brier, M., Ferree, T., Maguire, M., Moore, P., Spence, J., Tillman, G., Hart, J., Kraut, M. Frontal Theta and Alpha Power and Coherence Changes are Modulated by Semantic Complexity in Go/NoGo Tasks. *International Journal of Psychophysiology*, in press.

### **CONCLUSION:**

In summary, the assimilation of this unique multimodal human cognitive-motor assessment laboratory is a significant accomplishment and will allow for numerous innovative studies. As a result of this laboratory and work, we have successful in obtaining a 3-yearCDMRP DMRDP grant entitled “Novel Treatment of Emotional Dysfunction in PTSD,” using rTMS in conjunction with Cognitive Processing Therapy.

### **REFERENCES:**

1. Jafari, R. & Hart, J. Pupil Dilation in response to threat/non-threat pictures with varying luminance. Technical Report.
2. Ferree, T., Brier, M., Hart, J., Kraut, M. Space-time frequency analysis of EEG data using within-subject statistical tests followed by sequential PCA. *Neuroimage*, 45(1):109-21, 2009.
3. Maguire, M., Brier, M., Moore, P., Ferree, T., Ray, D., Mostofsky, S., Hart, J., Kraut, M. The influence of perceptual and semantic categorization on inhibitory processing as measured by the N2-P3 response. *Brain and Cognition*, 71, 59, 2009.
4. Brier, M., Ferree, T., Maguire, M., Moore, P., Spence, J., Tillman, G., Hart, J., Kraut, M. Frontal Theta and Alpha Power and Coherence Changes are Modulated by Semantic Complexity in Go/NoGo Tasks. *International Journal of Psychophysiology*, in press.

## Appendices

	<u>Page</u>
Jafari & Hart.....	8
Ferree et al.....	13
Maguire et al.....	26
Brier et al.....	34

# **Pupil Dilation in response to threat/non-threat pictures with varying luminance**

R. Jafari & J. Hart

*Abstract* - The goal of this study was to detect the level of arousal induced by pictures of negative, positive and neutral affective valence by measuring the pupil dilation of 4 subjects. In order to detect arousal from pupil dilation data, it was necessary to reduce the effect of luminance. This required comparing the picture luminance measures to the collected pupil dilation data, and applying additional processing to reduce the effect of luminance. The applied algorithm extracted the component of pupil dilation orthogonal to picture luminance measures. Although this processing increased the arousal “score” of some high arousal (as opposed to neutral) pictures, neutral pictures still had higher scores than expected. This suggests a nonlinear relationship between pupil dilation and luminance.

## **I. Introduction**

In our experiment, four subjects were exposed to threat, non-threat and scrambled pictures. Of these, 20 pictures were high arousal and positive valence (non-threat), 20 pictures were high arousal and negative valence (threat), and 40 pictures were low arousal, scrambled versions of the high arousal pictures. Each picture was displayed for 2 seconds, with an 8 second inter-stimulus interval.

Throughout the experiment, pupil dilation was measured with the iView X eye tracking system. Although pupil dilation was expected to be a function of arousal induced by pictures, luminance was also a factor in the overall pupil dilation. Previous studies have controlled for luminance in their stimuli in various ways. [1] used sound stimuli from the International Affective Digitized Sounds (IADS) set, to avoid the complications of visual stimuli. [2] and [3] used white text on a black background to keep luminance relatively constant. [4] converted International Affective Picture System (IAPS) pictures to grayscale pictures, and equated the average luminance of each class of pictures.

In this experiment, the pupil dilation results are adjusted to account for luminance. The mean luminance of each picture can be calculated from RGB values by converting to the Y component of the CIE color space [5]. The conversion from sRGB (color space convention used in MATLAB) to the luminance, Y, is given by [6].

## **II. Processing pupil diameter data**

Pupil dilation was determined by taking the difference between the end and start points of the 2 second stimulus duration. Invalid data due to blinking and transient noise were replaced using linear interpolation. Invalid data segments were detected by limiting the valid range of the mean and



standard deviation over an 80 ms period. Pupil dilation results were recorded over all 80 pictures for all 4 subjects.

### III. Adjusting for luminance

Because our goal was to find the pictures of highest arousal across all subjects, pupil dilation results had to be averaged across the 4 subjects. However, the mean and standard deviation of dilation with respect to a subject does not contribute any information to discriminate between pictures. The pupil dilation data for each subject was therefore converted to a zero-mean vector, and all four vectors were normalized to have the same standard deviation. Once the zero-mean, normalized vectors were averaged across the 4 subjects, the resulting vector was transformed to reduce the effect of luminance.

Consider the pupil dilation vector,  $\mathbf{p}$ , and the luminance vector,  $\mathbf{l}$ , of equal size.  $\mathbf{p}$  can be decomposed into components parallel and orthogonal to  $\mathbf{l}$ .

$\mathbf{p}$  can therefore be reduced to its component orthogonal to  $\mathbf{l}$ ,  $\mathbf{p}_{\perp}$ , by the following operation.

$$\mathbf{p}_{\perp} = \mathbf{p} - \frac{\mathbf{p} \cdot \mathbf{l}}{\mathbf{l} \cdot \mathbf{l}} \mathbf{l}$$

The result of this operation is displayed in Table 1.

**Table 1. First 20 pictures arranged in descending order according to**

	Name	Valence		
74	tgun_in_belt	-1	0.115347505	0.152155
52	apachecopter	-1	0.034490611	0.151945
79	nhot_chocolatescr	0	-0.0010266	0.145671
64	ncake_flower_scr	0	0.15235485	0.117463
38	troaring_tiger	-1	0.116284019	0.115561
51	twolfscr	0	0.121940414	0.111647
54	tfunnel_cloudscr	0	0.139537803	0.107237
62	smokehelijeepscr	0	0.120158398	0.105061
13	gunmanviewscr	0	0.009563267	0.093799
6	tsniper_rifle	-1	0.004372519	0.092999
40	aussie13scr	0	0.027024582	0.089859
63	n_beach_dog_n_baby	1	0.042247151	0.089525
14	nkittensscr	0	0.0308066	0.083023
69	nhorsescr	0	0.027850034	0.076117
10	nnice_brown_dogscr	0	0.06639312	0.071448

33	tblack-neck_snakescr	0	-0.05280666	0.052838
65	n_beach_dog__n_baby_scr	0	-0.01441656	0.052248
70	tsnowstorm	-1	-0.01706497	0.050081
24	ttwisterscr	0	0.099689556	0.047187
78	ngrand_parent's_facesscr	0	0.059495774	0.045928

Table 1 shows the 20 pictures corresponding to the largest elements in . The elements of can therefore be considered to be scores used to rank pictures according to arousal. The valence column indicates whether the picture has positive (1), negative (-1) or neutral (0) valence. Valence scores of -1 and 1 were expected to correspond to higher arousal scores while valence scores of 0 were expected to correspond to low arousal scores. The results in Table 1 indicate that the effect of luminance on pupil dilation may be significantly nonlinear, limiting the effectiveness of , or that only a few pictures resulted in significant arousal.

To compare the values of to the original pupil dilation vector, , Table 2 lists the 20 largest elements in .

**Table 2. First 20 pictures arranged in descending order according to**

	Name	Valence		
64	ncake_flower_scr	0	0.15235485	0.117463
54	tfunnel_cloudscr	0	0.139537803	0.107237
51	twolfscr	0	0.121940414	0.111647
62	smokehelijeepscr	0	0.120158398	0.105061
38	troaring_tiger	-1	0.116284019	0.115561
74	tgun_in_belt	-1	0.115347505	0.152155
67	tsuicide_gunman	-1	0.105319183	0.004066
47	tsuicide_gunmanscr	0	0.101810676	0.029326
24	ttwisterscr	0	0.099689556	0.047187
1	nelephants	1	0.092348936	0.01178
30	ttwister	-1	0.090130234	0.014674
37	ntomatoesscr	0	0.086980004	0.004621
73	thand_pointing_gun	-1	0.085774709	-0.03198
4	tdemon_dog	-1	0.074187284	-0.05516
2	nelephantsscr	0	0.071106304	-0.00039
50	npopsicles	1	0.067395423	0.026409
10	nnice_brown_dogscr	0	0.06639312	0.071448
39	n3fooddishes	1	0.063894199	-0.00389
78	ngrand_parent's_facesscr	0	0.059495774	0.045928
46	twolf	-1	0.058341368	0.013317

#### IV. Alternative techniques to estimate and reduce the effect of luminance

Alternate algorithms that were applied to the data set did not yield improved results, but may be useful if applied to larger data sets.

##### A. Alternative luminance estimate

Some pictures that may have caused high arousal corresponded to low values in . However, because subjects do not fixate their gaze on the background, the luminance of these pictures may be underestimated.

An alternate estimate of luminance is therefore the maximum mean luminance of any window in the picture. These luminance measures were calculated for and windows, but were found to have a lower magnitude of correlation with pupil dilation than overall luminance.

**Table 3. Correlation coefficient of different luminance measures and pupil dilation.**

Overall luminance	-0.741914564
Maximum window	-0.552441567
Maximum window	-0.388838877

Another problem with this measure is that it favors the non-scrambled pictures. The neutral, scrambled pictures have a relatively constant luminance throughout a picture, and therefore have smaller maxima.

A better approach may therefore be to apply a nonlinear function (such as log) to the overall luminance calculation because this may better describe the effect of luminance on pupil dilation.

##### B. Calculating for each subject

It is reasonable to assume that the effect of luminance on pupil dilation is more consistent than the subjects' affective response to pictures. Whenever a set of signals are averaged, the resulting signal will emphasize factors common to all signals, while reducing factors that vary between signals. We can therefore expect that the effect of luminance will increase if pupil dilation vectors are averaged. This hypothesis is supported by Table 4, which shows correlation coefficients between various pupil dilation vectors and the luminance vector, which is constant between subjects. Specifically, it compares the correlation coefficients calculated from the dilation vectors of individual subjects, to the correlation coefficients calculated from the average dilation vector across all subjects.

**Table 4. Correlation coefficients calculated by correlating pupil dilation vectors to a constant luminance vector.**

Subject 0	-0.53574
Subject 1	-0.71699
Subject 2	-0.45791
Subject 3	-0.61908
Average dilation vector across subjects	-0.74191

From Table 4 it is evident that luminance becomes a more significant factor if the dilation vectors are averaged across subjects. To reduce the effect of luminance it may therefore be beneficial to calculate separately for each subject before finding the average vector across subjects. Although this technique yielded results similar to Table 1, it may prove superior on a different data set.

#### REFERENCES

- [1] T. Partala and V. Surakka, "Pupil size variation as an indication of affective processing," *International Journal of Human-Computer Studies*, vol. 59, pp. 185-198, July 2003.
- [2] G. J. Siegle, E. Granholm, R. E. Ingram, and G. E. Matt. "Pupillary and reaction time measures of sustained processing of negative information in depression," *Biological Psychiatry*, vol. 49, no. 7, pp. 624-636. April 2004.
- [3] J. S. Silk et al., "Pupillary reactivity to emotional information in child and adolescent depression: links to clinical and ecological measures," *Am. J. of Pscychiatry*, vol. 164, pp. 1873-1880, Dec. 2007.
- [4] M. M. Bradley, L. Miccoli, M. A. Escrig, and P. J. Lang, "The pupil as a measure of emotional arousal and autonomic activation," *Psychophysiology*, vol. 45, no. 4, pp. 602-604, July 2008.
- [5] J. Schanda, *Colorimetry: Understanding the CIE System*. Hoboken, New Jersey: John Wiley, 2007, pp. 29-30.
- [6] M. Stokes, M. Anderson, S. Chandrasekar, R. Motta, "A standard default color space for the internet: sRGB," *International Color Consortium*, Nov. 1996. [Online]. Available: <http://www.color.org/sRGB.xalter>.



Contents lists available at ScienceDirect

NeuroImage

journal homepage: [www.elsevier.com/locate/ynimg](http://www.elsevier.com/locate/ynimg)

## Space–time–frequency analysis of EEG data using within-subject statistical tests followed by sequential PCA

Thomas C. Ferree<sup>a,\*</sup>, Matthew R. Brier<sup>b</sup>, John Hart Jr<sup>b,c</sup>, Michael A. Kraut<sup>d</sup>

<sup>a</sup> Department of Radiology, University of Texas Southwestern Medical Center, Dallas, Texas, TX 75390-8896, USA

<sup>b</sup> Center for Brain Health, School of Behavioral and Brain Sciences, University of Texas at Dallas, USA

<sup>c</sup> Department of Neurology, University of Texas Southwestern Medical Center, Dallas, Texas, USA

<sup>d</sup> Department of Radiology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

### ARTICLE INFO

#### Article history:

Received 7 May 2008

Revised 7 August 2008

Accepted 5 September 2008

Available online xxxx

### ABSTRACT

A new method is developed for analyzing the time-varying spectral content of EEG data collected in cognitive tasks. The goal is to extract and summarize the most salient features of numerical results, which span space, time, frequency, task conditions, and multiple subjects. Direct generalization of an established approach for analyzing event-related potentials, which uses sequential PCA followed by ANOVA to test for differences between conditions across subjects, gave unacceptable results. The new method, termed STAT-PCA, advocates statistical testing for differences between conditions within single subjects, followed by sequential PCA across subjects. In contrast to PCA-ANOVA, it is demonstrated that STAT-PCA gives results which: 1) isolate task-related spectral changes, 2) are insensitive to the precise definition of baseline power, 3) are stable under deletion of a random subject, and 4) are interpretable in terms of the group-averaged power. Furthermore, STAT-PCA permits the detection of activity that is not only different between conditions, but also common to both conditions, providing a complete yet parsimonious view of the data. It is concluded that STAT-PCA is well suited for analyzing the time-varying spectral content of EEG during cognitive tasks.

© 2008 Elsevier Inc. All rights reserved.

### Introduction

Cognitive experiments involve stimuli delivered to the subject, and responses generated by the subject, in time frames that depend upon the task. Brain responses that involve increased synchrony with a consistent phase relationship to an event, e.g., stimulus or response, contribute to the event-related potential (ERP). Brain responses that involve decreased synchrony, or increased synchrony without a consistent phase relationship to an event, are best detected with time–frequency analysis. In cognitive tasks with long duration, time–frequency analysis is expected to be more fruitful than ERP analysis. While time–frequency analysis is relatively straightforward, a considerable challenge remains to reduce and summarize the numerical results, which span space, time, frequency, task conditions, and subjects.

Principal component analysis (PCA) has been applied extensively to ERP analysis, in order to reduce the waveforms in spatial and temporal dimensions (Spencer et al., 1999, 2001). The ERP is computed in each condition, by averaging the post-stimulus time series across trials. The number of trials in each condition and the variance of the mean are not retained. Then PCA is applied sequentially to spatial and

temporal dimensions, and the resulting factor scores are submitted to ANOVA to test for differences between conditions. The group of subjects, rather than repeated trials, is used as the statistical ensemble when testing for differences between conditions. This approach, which we call PCA-ANOVA, is reported to work well for ERP analysis, and is now used widely. Its main limitations are that PCA-ANOVA can test only for differences between conditions, and requires multiple subjects on which to base these tests.

When extending to time–frequency analysis, the needs for data reduction are even greater, and the very nature of the data is different. Only a few studies have used PCA together with time–frequency analysis. Bernat et al. (2005) combined time–frequency analysis with PCA, but their emphasis was a comparison between different methods of time–frequency analysis. Tenke and Kayser (2005) studied the effects of transforming the power spectrum, and using an explicit reference versus the surface Laplacian. Our findings support this previous work, but neither group addressed the key question of how best to integrate PCA with statistical testing.

Following the approach established in the ERP literature, we applied PCA sequentially to frequency, space, and time dimensions, then submitted the resulting scores to ANOVA to test for differences between conditions. The results were found to be unstable, and changing the order of dimensions did not resolve the problems. We suspected that PCA, when applied first, was unable to isolate task-related changes, because the power spectrum is dominated by

\* Corresponding author.

E-mail address: [tom.ferree@gmail.com](mailto:tom.ferree@gmail.com) (T.C. Ferree).

features that may not be task related. If the task-related effects do not contribute the greatest variance to the matrix passed to PCA, then PCA will not isolate the task-related activity in the highest components. Because Varimax rotation behaves poorly with many components, and because the overall goal is data reduction, it is important to organize the analysis so that task-related activity appears in the first few principal components.

We hypothesized that a better approach would put statistical testing at the beginning of the analysis, in order to isolate task-related variance in single subjects. Statistical testing for differences between power spectra is standard in a large body of work on event-related synchronization (ERS) and de-synchronization (ERD), in which these tests are conducted in single subjects (Pfurtscheller and Lopes da Silva, 1999; Delorme and Makeig, 2004). The number of trials in each condition, and the variance in the estimate of the mean, are used to test statistical significance. For visualization purposes, differences that are not statistically significant are often rounded to zero. In the new approach called STAT-PCA, we tested for differences between conditions in single subjects, then followed with PCA for data reduction, and found that results were highly stable.

Within-subject statistical testing also solves several other problems inherent in the analysis of cognitive data with PCA. First, it permits testing for differences not only between conditions, but also between a given condition and baseline. We use this fact to reveal activity that is common to both conditions. Second, it has been noted that the rotation ambiguity of PCA factors may result in misallocation of variance (Wood and McCarthy, 1984), giving linear combinations of activity in the two conditions (Dien, 1998). By conducting statistical tests within subjects, activity that is different between conditions, and activity that is common to both conditions, are separated before decomposing with PCA, so an important caveat of PCA is eliminated. Third, statistical testing in single subjects isolates task-related activity in single subjects, and this facilitates clinical diagnosis in which single subjects are the focus of investigation.

## Methods

### Participants

The subjects were 25 young adults between the ages of 18 and 29. All were right-handed, and 10 were male. All were free from neurological or psychiatric disorders by self-report. Written informed consent was obtained from each subject prior to testing. This study was conducted according to the Good Clinical Practice Guidelines, the Declaration of Helsinki, and the U.S. Code of Federal Regulations. Written and informed consent was obtained from all participants according to the rules of the Institutional Review Board of The University of Texas at Dallas.

### Stimuli and task

The data set upon which we developed these methods is part of a continuing study of semantic memory retrieval (Slotnick et al., 2002; Assaf et al., 2006). The stimuli consisted of pairs of written words, which consisted of features of familiar objects (e.g., 'desert' and 'humps'). The subjects were to determine whether the two features combined to result in retrieving the memory of a specific object (e.g., 'camel'). Other word pairs were chosen to lead to no retrieval (e.g., 'mane' and 'wings'). Subjects were instructed that the target needed to be a specific object, not merely an association between the two words. Fifty trials comprised stimulus pairs that have been shown in previous work (Assaf et al., 2006; Brier et al., 2008) to elicit retrieval of a specific object, and 50 were non-retrieval trials. The same feature words used in the object retrieval pairs were used in the non-retrieval pairs, but were re-paired with a semantically unrelated word. Each word pair appeared on the screen for 3 s, separated by a blank screen that

appeared for 2–3 s (randomized). While it is true that the researchers defined which word pairs should or should not elicit retrieval, the small number of 'incorrect' responses indicates that this task does access semantic memory. The few trials with incorrect responses were also discarded.

### Data acquisition

EEG data were acquired with a 64-channel, Synamps II system (Compumedics, Inc.). Data were sampled at 1 kHz and hardware filtered at 200 Hz. Electrode impedances were typically below 5 k $\Omega$ , although some were slightly higher. An experienced EEG technician preprocessed the data manually. First, data recorded from poorly functioning electrodes were identified visually and removed. Second, eye blink artifacts were removed by a spatial filtering algorithm in the Neuroscan Edit software (Compumedics, Inc.), using the option to preserve the background EEG. This option uses the singular value decomposition of a 'clean' data segment to optimize the removal of eye blinks from the continuous data (M. Pflieger, personal communication). Third, time segments containing significant muscle artifacts or eye movements were rejected.

During acquisition, time-locking events were placed in the EEG record corresponding to the white computer screen, the onset of word-pair stimuli, and button-press responses of two types. For spectral analysis, a baseline interval was defined as 1 s prior to the stimulus. In order to study stimulus-related activity, a peri-stimulus interval was defined from –1 to 3 s. The data were epoched accordingly and exported to Matlab (Mathworks, Inc.) for further analysis. We have begun to explore the benefits of epoching relative to responses, but for brevity only peri-stimulus results are reported here.

### Reference correction

The data were recorded with a reference electrode located near the vertex, which results in small amplitudes over the top of the head. In order to correct for this effect, the data were re-referenced to the average voltage at each time point, which approximates the voltage relative to infinity (Nunez, 1981). In order to minimize a known bias in the electrode-based average reference (Junghofer et al., 1999), a spline-based estimate of the average scalp potential (Ferree, 2006) was computed using spherical splines (Perrin et al., 1989). Placing the electrode cap on a realistic phantom head, the electrode coordinates were digitized (Polhemus, Inc.), and these coordinates were used to fit the splines for each subject. The integrity of the spline interpolation was confirmed visually, by comparing waveforms of arbitrarily deleted channels with the original waveforms in those channels. The integrity of the spline-based average reference was confirmed by comparing topographic maps of baseline alpha power with similar maps using the cap reference, and the electrode-based average reference. In subjects with a small number of bad electrodes, the splines were used to interpolate those electrodes, to yield a total of 62 data channels in every subject. Ensuring the same number of electrodes in all subjects facilitates the matrix manipulations in sequential PCA.

### Time–frequency analysis

Throughout the peri-stimulus interval, time-dependent Fourier power spectra were estimated in 0.5-second wide windows, moving in 0.05-second steps. The time of each window was defined as the center of the nonzero data in that window. The earliest time was –0.75 s, and the latest time was 2.75 s, because the centers of 0.5-sec windows cannot reach the ends of the epoch. Fourier power spectra were computed using the *pwelch* function implemented in Matlab (Mathworks, Inc.). In each window, the time series was linearly

detrended to reduce spectral leakage from the zero-frequency bin, cosine tapered to reduce spectral leakage generally, and zero-padded to 1-s duration to achieve 1-Hz frequency resolution. Each time series was then Fourier transformed, magnitude squared, and suitably normalized to obtain the power spectral density (PSD) in units  $\mu V^2$ /Hz. For each condition, the result was averaged across all trials, to obtain the best statistical estimate of the PSD.

The time-averaged PSD in the baseline interval was computed two ways for comparison. In both ways, the PSD was averaged across all trials in both conditions. In the first way, the time averaging was accomplished using the Welch method, in which the 1-s baseline interval was divided into three 0.5-s windows with 50% overlap (Welch, 1967). In the second way, the time averaging was accomplished by averaging the time-varying PSD over all time points prior to the stimulus. This was done because we observed in several subjects that the power values in the Welch windows were not always representative, occasionally missing large fluctuations and leading to inaccurate estimates of the temporal mean. The relationships between the estimates of baseline power, and the effects on PCA-ANOVA and STAT-PCA, are discussed in the Results.

If the results of our analysis are to be used to explain cognitive processes in terms of neural oscillations, we should state clearly what oscillations are included in our analysis. In conventional terminology, 'evoked' activity has a precise time and phase relationship with a temporal event, while 'induced' activity does not. The ERP isolates evoked activity, but spectral analysis is needed to detect induced activity. Because evoked activity also contributes to spectral power, a direct application of spectral analysis must be considered to reflect both evoked and induced activity. Some researchers have attempted to isolate induced activity, by subtracting the ERP prior to time-frequency analysis (Kalcher and Pfurtscheller, 1995; Ding et al., 2000; Truccolo et al., 2002). In the present work, it is not our goal to distinguish evoked and induced activity, and we have not implemented any method to subtract the ERP. Our results must therefore be interpreted to include both evoked and induced activity. Because it is difficult to maintain phase locking to the stimulus for long times, however, we expect that later times are dominated by induced activity, especially at higher frequencies.

#### *Difference-mode and common-mode responses*

The standard PCA-ANOVA approach, applied to ERP waveforms, is usually arranged to test for differences between conditions (Spencer et al., 1999, 2001; Dien et al., 2003). The calculation of the ERP waveform in each condition begins with subtracting the baseline, defined as the average of the pre-stimulus time series across time and trials. Baseline subtraction is required in ERP analyses, because EEG time series are prone to slow drift from electrode polarization and imperfect amplifier properties. Yet activity in the baseline interval may include residual activity from the previous response and/or anticipation of the next stimulus. For this reason and others, many ERP researchers prefer to focus only on differences between conditions.

In the ERD/ERS literature, it is common to quantify brain oscillations relative to baseline (Pfurtscheller and Lopes da Silva, 1999; Delorme and Makeig, 2004). In the power spectrum, any electrode or amplifier drift appears in the zero-frequency term, so this effect is not central. (An exception is that any power in the lowest frequency bin, e.g., less than 0.5 Hz when using 1-Hz frequency bins, may contaminate higher frequencies by spectral leakage, but this effect is minimized by detrending the data in each window before Fourier transforming.) Of course, the baseline interval may still be contaminated, by the previous response or anticipation of the next stimulus. Because the Fourier transform is squared in each trial to compute power, and because the moving windows have non-zero width determined by the taper, the same variation in the inter-stimulus interval may not be as effective.

Despite the issues inherent to defining and interpreting a baseline interval, we argue that it is compelling to consider *both* differences between conditions *and* differences relative to baseline in parallel. In the case of our semantic retrieval task, for example, it seems likely that studying the words visually, searching for associations, and launching a motor response are common to both conditions, while retrieving a word from semantic memory occurs only in one condition. Although the mental processes of semantic retrieval are not yet understood well enough to list and categorize each one, considering both kinds of activity obviously gives a more complete picture of the data. Consider a hypothetical example, in which there is greater power in condition A than condition B, for some frequency band and time interval. From this information alone, any of the following could be true: 1) condition A has ERS that condition B does not, 2) condition B has ERD that condition A does not, 3) conditions A and B both have ERS, but condition A has more, 4) conditions A and B both have ERD, but condition B has more. Only by studying the differences relative to baseline is it possible to distinguish these cases.

Following this logic, we define the 'difference-mode' response as the difference of moving-window power spectra between two conditions, without reference to baseline, and define the 'common-mode' response as the difference of moving-window power spectra between *both* conditions and baseline. By *both* conditions, we mean the average of the PSD in both conditions, weighted by the number of trials in each condition; that is equivalent to taking the union of the trials in both conditions before averaging across trials. The rationale for pooling the responses from both conditions to form the common-mode response, rather than looking at each response separately relative to baseline, is that half of the information in the separate responses is already contained in the difference-mode response, and the common-mode response is orthogonal to that. Taken together, therefore, the difference-mode and common-mode responses give the most complete yet parsimonious view of task-related activity. We submit the results from both modes separately to PCA. Once the prominent factors from both modes are identified, these can guide inference about the behavior in each condition.

To be precise, consider the algebraic definitions of common mode (CM) and difference mode (DM), assuming the number of trials in the two conditions is equal. Let  $R$  represent the time-varying PSD in the retrieval condition,  $N$  represent the time-varying PSD in the non-retrieval condition, and  $B$  represent the time-averaged PSD in the baseline interval. At each frequency, electrode, and time point, we have  $DM = R - N$ , and  $CM = (R + N)/2 - B$ . Solving for  $R$  and  $N$  gives  $R - B = CM + DM/2$ , and  $N - B = CM - DM/2$ . Thus the response in each condition relative to baseline may be obtained trivially from the common-mode and difference-mode results. In the present work, difference-mode STAT-PCA permits a direct comparison with PCA-ANOVA, and common-mode STAT-PCA gives new information not accessible with PCA-ANOVA.

#### *Statistical tests for differences between power spectra*

In most ERP studies, only the within-subject average response is carried forward, e.g., to PCA-ANOVA, and neither the number of trials nor the variance across trials is used. Instead, the statistical ensemble used to test for differences between conditions is comprised of many subjects in the study. In our initial attempts, we adopted this approach, submitting the trial-averaged, time-varying power in each condition to three-way PCA-ANOVA. Empirically, the results were unsatisfactory (see Results). Theoretically, a disadvantage of conducting PCA first is that statistical variability in the estimates of the power spectrum, which are inherent for stochastic signals, might overly influence the PCA. On these grounds, we hypothesized that within-subject statistical tests for differences in power spectra, either between conditions or relative to baseline, would improve the results.



Statistical testing for differences between power spectra is a reasonably well-developed topic. In its simplest form, the log of the power spectrum estimate is assumed to be Gaussian distributed. The estimate based upon a finite number of samples is biased, however, by an amount that depends upon the number of samples (Thomson and Chave, 1991; Bokil et al., 2007). Using analytic expressions in these references, we corrected for this bias and estimated the variance in the mean using the appropriate number of samples in each case. The bias correction is a simple function of the number of samples, which is subtracted from the original calculation of the PSD. In the post-stimulus interval, the number of samples contributing to the PSD in each condition was taken to be twice the number of trials in that condition, accounting for sine and cosine contributions. (In the calculation of baseline power and the common mode, the number of trials was equal to the sum of the trials in both conditions.) Furthermore, in the baseline interval, because the 50% overlap combined with cosine tapering leads to approximately independent samples, the number of samples was tripled, corresponding to three 0.5-s windows in the 1-s baseline interval.

The PSD is computed at many electrodes, frequencies, and time points. By setting  $\alpha=0.05$ , false positives are expected at this rate. Correcting for multiple comparisons is non-trivial when the data are correlated. In the spatial dimension, data at different electrodes are correlated due to volume conduction. In the spectral dimension, spectral leakage causes neighboring frequency bins to be correlated. In the temporal dimension, overlapping windows cause neighboring time points to be correlated. One remedy in the spectral domain is to keep only contiguous sets of frequencies that are wider than the bandwidth of the analysis (Bokil et al., 2007), and another is to use the false-discovery rate (Durka et al., 2004), but neither of these approaches as published addresses false-positives in the full three dimensions of space–time–frequency. In the present work, we have chosen for simplicity not to correct for multiple comparisons. If false positives occur randomly, they should have little impact on the PCA analysis applied to multiple subjects. Indeed, the results below support the assertion that only meaningful, task-related activity emerges from STAT-PCA.

#### *Input to principal component analysis*

This work contrasts two approaches to integrating time–frequency analysis with statistical tests. In the standard method, PCA-ANOVA, the time-dependent power values in each condition are input to sequential PCA (see below) and the resulting scores are passed to ANOVA. This is a one-way (condition) repeated measures ANOVA with two levels (retrievals and non-retrievals). In this approach, the variance across trials is not retained, and the statistical ensemble used in the ANOVA is comprised of multiple subjects. In the new method, STAT-PCA, statistical tests between conditions are conducted in each subject and electrode separately, and the statistical ensemble is comprised of multiple trials. Insignificant differences are rounded to zero, in effect, eliminating noise from the results. Only the non-zero spectral differences are passed to PCA for data reduction.

Although the statistical tests for differences between spectra in STAT-PCA assumed that the log-power spectrum is Gaussian distributed, we used PSD in units of  $\mu\text{V}/\text{Hz}^2$  not dB as input to PCA. This choice does not affect the set of frequencies that show significant differences, but does affect the numerical values submitted to PCA. We adopted this method after we tried both  $\mu\text{V}/\text{Hz}^2$  and dB units and the former gave much better results. The log transformation renders large power values not so different from the small power values, and this had several adverse effects on the PCA results: many more factors retained, less distinctive factor loadings, and poorer agreement with group averaged data. We therefore kept the units of  $\mu\text{V}/\text{Hz}^2$  as input to PCA, consistent with the recommendations of Tenke and Kayser (2005), but set the differences to zero according to the aforemen-

tioned statistical test. Even with this choice, which tends to under-emphasize the power at higher frequencies, we obtained differences in the 20–35 Hz range that were reported previously in this task (Slotnick et al., 2002).

#### *Principal component analysis*

In our description of PCA, we use standard terminology and matrix organization, with some small exceptions. The data matrix is arranged with columns indexing variables and rows indexing samples. Following standard conventions in PCA analysis, the column-means are subtracted, i.e., in each column separately the mean across rows is subtracted. Following the consensus in the ERP literature (Kayser and Tenke, 2003; Dien et al., 2005), we use the covariance matrix rather than the correlation matrix. With this convention, PCA is equivalent to singular value decomposition (SVD): the right eigenvectors are called the component or factor 'loadings', and the left eigenvectors times the singular values are called the component or factor 'scores'.

In ERP analysis, PCA has been applied sequentially to reduce the results in the spatial and temporal dimensions. An important consideration is the order in which PCA is applied to the various dimensions. An early work applied PCA to the spatial dimension (Donchin, 1966), and later works applied PCA to the temporal dimension (Curry et al., 1983; Donchin and Heffley, 1979; Mocks and Verleger, 1991). More recent works applied PCA spatially then temporally (Spencer et al., 1999, 2001; Dien et al., 2003). The choice to apply spatial PCA before temporal PCA for ERP analysis was based on the argument that 'components are defined by unique patterns of scalp distributions' (Spencer et al., 2001). It has also been suggested to apply temporal PCA first (Dien and Frishkoff, 2005), because spatial components may overlap due to volume conduction. Despite all these well-reasoned arguments, no study has yet compared the effects of order in sequential PCA.

In our initial investigations, we studied PCA-ANOVA and STAT-PCA for all six possible orderings of frequency, space, and time. In order to keep this report tractable, we focus on one order. We arrived at this order by following one of the earlier arguments, that the best order is determined by the inherent separability of the data (Dien, 1998). First, cognitive processes are accompanied by oscillations in characteristic frequency bands. Bands have nonzero width, but do not typically overlap. This separability suggests that spectral PCA should be conducted first. Second, time–frequency analysis involves moving windows with some non-zero width, and this blurs time resolution below that of ERP analysis. This suggests that temporal PCA should be conducted last. On these arguments, we suggest that a sensible starting point is spectral–spatial–temporal PCA. Based upon our initial investigations, which spanned all six possible orderings, our impression is that this ordering produced among the most stable and sensible results for our data set.

To perform sequential PCA in this order, the data are arranged into a matrix, with columns indexing frequencies, and rows indexing the result of concatenating electrodes, time-points, conditions (PCA-ANOVA only), and subjects. First, PCA is applied to obtain the spectral loadings, and the largest factors are retained. For each spectral factor retained, the corresponding factor score is reshaped to form a matrix with columns indexing electrodes, and rows indexing the concatenation of time points, conditions (PCA-ANOVA only) and subjects. Second, PCA is applied to obtain the spatial loadings, and the largest factors are retained. For each spatial factor retained, the corresponding score is reshaped to form a matrix with columns indexing time points, and rows indexing the concatenation of conditions (PCA-ANOVA only) and subjects. Third, PCA is applied to obtain the temporal loadings, and the largest factors are retained. In PCA-ANOVA, the temporal scores are submitted to ANOVA. In STAT-PCA, the temporal scores are simply kept as the final result.



### Factor retention

The main objective of PCA is dimension reduction, but determining the precise number of components or factors to retain is notoriously difficult (Zwick and Velicer, 1986; Hayton et al., 2004). All of the standard methods are based upon the eigenvalues. The goal is to identify a small set that capture most of the data variance, and are distinct from the remaining set of smaller eigenvalues. In the simplest method, the eigenvalues are plotted in a 'scree' plot, and the cut point is determined by eye (Cattell, 1966). Despite its wide use, this method is highly subjective, especially when the eigenvalues decrease gradually, or when there are several distinct steps. In the present data, we often have one very large eigenvalue, followed by one or more visible steps. To make the choice less subjective, we implemented two statistical techniques that should bracket the best choice.

The first technique is the maximum profile likelihood (MPL), which aims to determine the cut point that leads to the most natural grouping of eigenvalues (Zhu and Ghodsi, 2006). After exploring this technique extensively, we have arrived at the following impressions. When a few eigenvalues are large and similar, standing well above the others, MPL picks these few. When the first eigenvalue is much larger than the others, MPL tends to pick it, even if there is a second elbow in the scree plot just a few points away. In this way, MPL appears either to perform well, or to underestimate the number of factors. MPL has the advantage of being very efficient computationally.

The second technique is parallel analysis (PA), which is based upon rejecting the null hypothesis that the eigenvalues are the same as those of a random matrix with the same dimensions and distribution of values (Horn, 1965). Comparison studies using model data agree that this is the most reliable technique (Zwick and Velicer, 1986; Hayton et al., 2004), although it is not widely used. One study showed that, when PA is wrong, it tends to overestimate the number of factors to retain (Zwick and Velicer, 1986), although another study points out that PA tends to underestimate the number of factors when the first eigenvalue is large (Turner, 1998). Despite having a single large eigenvalue in our data, we have found that PA always estimates more factors than MPL.

In order to generate random matrices for PA, we shuffled the values in the original matrix and computed the eigenvalues using identical procedures. For each eigenvalue generated from the null distribution, we computed the mean across the 100 random matrices. Eigenvalues from the real matrix that were greater than the mean eigenvalue from the random matrices were considered descriptive of the covariance structure of the data. The use of the mean has precedent (Hayton et al., 2004), but has also been criticized as setting the false-positive rate to 0.5. An obvious remedy is to set the false-positive rate to 0.05 (Glorfeld, 1995), but this requires many more random matrices, because evaluating the tails of a distribution is much more demanding computationally than evaluating the mean. Because PA is already quite demanding, and because a large false-positive rate corresponds to overestimating the number of factors, we used the mean of 100 random matrices as the threshold, and we interpret this PA estimate as an upper bound.

### Factor rotation and refinement of factor retention

PCA decomposes a data matrix into orthogonal components. While this is often effective for separating signal and noise subspaces, it is well known that the factors retained according to relative variance alone may not provide the most useful factorization of the signal. For this reason, it is normal to apply an additional transformation, such as factor rotation, to satisfy some additional constraint. In the present work, we focus on Varimax rotation, which aims to simplify the structure of each factor by loading its variance onto the smallest number of elements. In applications to ERP data, there is general

agreement that Varimax rotation helps separate distinct cognitive components (Kayser and Tenke, 2003; Dien et al., 2005).

As described in the previous subsection, the eigenvalue-based techniques MPL and PA are helpful in determining the number of factors to retain, but they leave some ambiguity and subjectivity in the choice. In the present work, we have made inroads toward an improved method for factor retention. It is based upon the common understanding that retaining too many factors becomes problematic when using rotation (Dien et al., 2007). Our idea is to compare the rotated factor loadings with the data, as a function of the number of factors retained. In order to compare the factor loadings with the data, we need some measure of the data that has the same dimension as the factor loadings. Because the column-means are subtracted prior to PCA, and because PCA is fundamentally a variance-based technique, the simplest non-zero data measure is the column-variance. Analogous to the definition of column-mean (see above) the column-variance is the variance of each column using the rows as samples. Of course, the column-variance of the data matrix contains less information than the full covariance matrix, but still provides a sensible data measure for evaluating the rotated factors.

To apply this idea in practice, in each step of sequential PCA, we plotted the original and rotated eigenvectors along with the normalized column-variance (see Fig. 3). Varying the number of factors retained within the range bracketed by MPL and PA, we retained the minimum number of factors necessary to explain the prominent features in the data. This procedure was carried out to decide the number of factors to retain in each step of sequential PCA. The choice to limit the number of factors to the prominent features in the column-variance, within the plausible range bracketed by MPL and PA, amounts to a conservative choice of the number of factors, because we know that the full covariance matrix contains more information. This use of the eigenvectors in conjunction with the eigenvalues to make choices about factor retention deserves more development, but even in its present form it is evident that using the rotated eigenvectors in this way is more rigorous than using the eigenvalues alone.

### Factor visualization and interpretation

Decisions about factor retention define a *factor tree* that associates each spectral factor with one or more spatial factors, and each spatial factor with one or more temporal factors. Each set of spectral, spatial, and temporal factors defines a *factor triplet*. Spectral and temporal loadings are plotted as two-dimensional curves, and spatial loadings are plotted as topographic maps. The loadings are normalized, but the signs are arbitrary. In order to visualize the spectral and spatial loadings consistently, we flipped their sign if their maximum absolute value corresponded to a negative value. Most spectral loadings had a single peak, although sub-peaks are often present. Most spatial loadings were either focal or bi-focal maps, appearing to represent semi-localized oscillators. Most temporal loadings were less compact, and the choice of orientation was usually arbitrary. In PCA, the (column-mean subtracted) data matrix is approximated as a product of spectral, spatial, and temporal loadings. In our conventions, any negative signs were effectively assigned to the temporal loadings as follows.

In order to determine the most sensible sign for the temporal loadings, we compared them to the subject-averaged data. For each triplet, we selected the peak frequency of the spectral factor, and one or more peak electrodes of the spatial factor. For these peaks, we averaged the common-mode and difference-mode PSD values across subjects. The temporal factors were flipped and scaled to have maximum similarity with the subject-averaged PSD. Visualization of the temporal loading along with the group-averaged data confirms the internal validity of the process, and permits each factor to be interpreted as ERD versus ERS.

# Metric for component similarity

This paper demonstrates that STAT-PCA performs better than PCA-ANOVA in several ways, and showing that requires comparing PCA factors quantitatively. Each *factor triplet* is comprised of a spectral loading  $F$ , a spatial loading  $S$ , and a temporal loading  $T$ . For two triplets:  $(F_a, S_a, T_a)$  and  $(F_b, S_b, T_b)$ , we define the *triplet similarity metric*

$$\Gamma_{ab} = (F_a^T F_b) (S_a^T S_b) (T_a^T T_b)$$

where  $T$  denotes vector transpose.  $\Gamma_{ab}=1$  implies a perfect match, and  $\Gamma_{ab}=0$  implies perfect orthogonality. Because small variations in the data can change the ordering of factors with similar eigenvalues, it is important to consider all possible orderings. Consider the general case, in which set  $A$  has  $M$  triplets, and set  $B$  has  $N$  triplets. To compare all pairings of factors from sets  $A$  and  $B$ , we constructed the matrix  $\Gamma_{ab}$ , for  $a=1, \dots, N$  and  $b=1, \dots, M$ . We summarized this matrix by computing the maximum of each row, to identify for each factor in set  $A$  the most similar factor in set  $B$ , independent of order. In order to generate a summary statistic across subjects, we took the maximum along the largest dimension of  $\Gamma$  (e.g., if  $a > b$  then the maxima were taken across the rows). This vector was then averaged to obtain a single global metric for each subject. These values were subjected to a paired  $t$ -test to determine if STAT-PCA was in fact more stable than PCA-ANOVA.

## Results

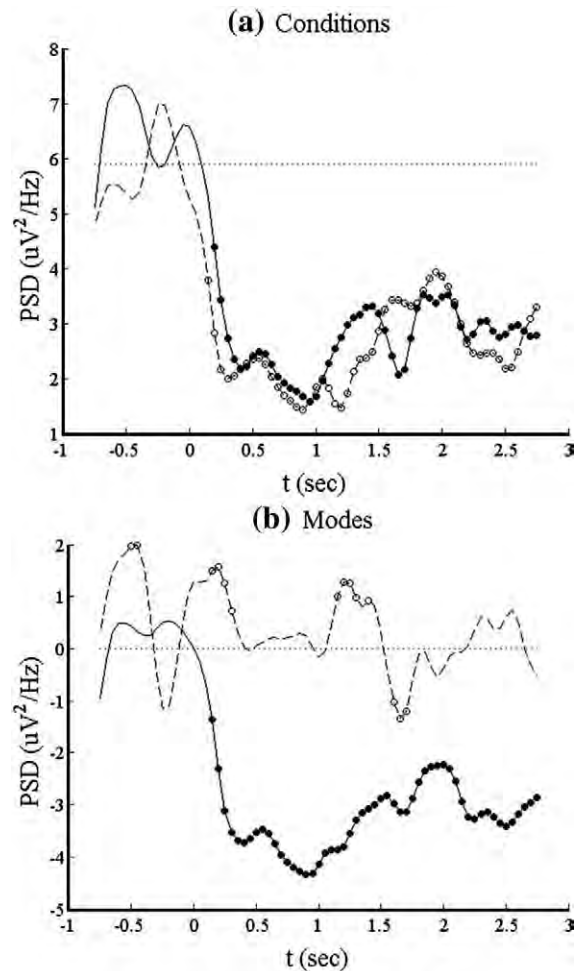
### Time-varying power and mode transformation in single subjects

Fig. 1a shows the time-varying power for a single subject, electrode OZ, frequency 10 Hz, relative to the Welch baseline power. Both conditions show power fluctuations during the baseline interval, decreased power after stimulus presentation, and a return toward baseline starting near 1 s. Fig. 1b shows the difference-mode (dashed) and common-mode (solid) responses. The common mode captures the strong decrease that is common to both conditions. The difference mode is much smaller, although some deflections are visible.

Fig. 1a also shows with dots the time points at which the power in each condition was significantly different from baseline ( $p < 0.05$ ). Neither condition was significantly different from baseline prior to the stimulus. Fig. 1b shows with dots the time points at which the power in each mode was significantly different from baseline ( $p < 0.05$ ). The common mode (solid dots) was significantly different from baseline only after the stimulus. The difference mode (open dots) showed significant differences sporadically, including some points prior to the stimulus. Because the stimuli were randomized, there can be no systematic difference in baseline between the two conditions. We conclude that the differences between conditions prior to the stimulus are due to random variations, and the points that passed the statistical test in this interval are false positives. Because false positives occur randomly across subjects, however, they are not expected to influence the PCA applied to multiple subjects.

### Results of automated tests for factor retention

Fig. 2 shows spectral eigenvalue plots computed for PCA-ANOVA and STAT-PCA in difference mode. The filled dots show MPL and PA recommendations for the number of retained factors. In all cases shown, MPL = 1 or 2, and PA > MPL. Fig. 2a was found for PCA-ANOVA without baseline subtraction. The second factor represents 60 Hz noise, as described below. Fig. 2b was found for PCA-ANOVA with baseline subtraction. Even though the MPL and PA values for factor retention in Fig. 2b are identical to those in Fig. 2a, baseline

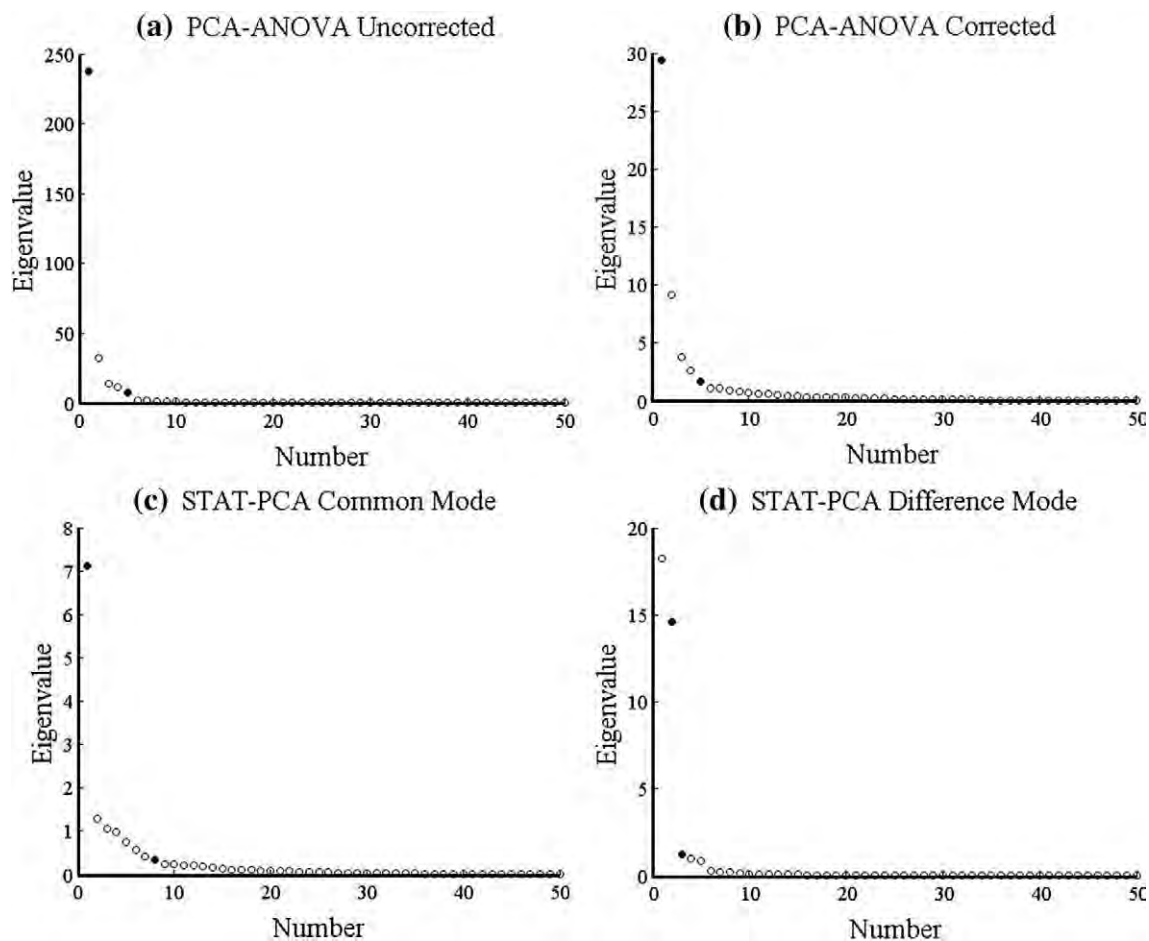


**Fig. 1.** Time-varying power at 10 Hz in electrode OZ for single subject: (a) retrieval condition (solid), non-retrieval condition (dashed), baseline (dotted); (b) difference mode (dashed), common mode (solid), zero (dotted). Dots show statistical significance for  $p < 0.05$ .

subtraction moves 60 Hz noise from the second factor to beyond the fifth factor. Fig. 2c was found for difference-mode STAT-PCA, and Fig. 2d was found for common-mode STAT-PCA. Broadly speaking, the eigenvalue plots for STAT-PCA are similar to PCA-ANOVA, featuring 1–2 large values, followed by a small set preceding an elbow in the range 5–8. The values for MPL and PA bracket any visible elbow, except in this example of common-mode STAT-PCA. Further exploration of factor retention choices in common-mode STAT-PCA showed that, although there appears to be a step after five factors, only two factors were interpretable.

### Refined factor retention using factor rotation

The MPL and PA algorithms provide useful guidelines for the number of factors to retain. In order to select a number within this range, we used additional information about how the rotated eigenvectors correspond with the column-variance of the data matrix. Fig. 3 shows the un-rotated (thin dashed lines) and rotated (thin solid lines) along with the column-variance (thick solid line), for difference-mode spectral STAT-PCA, and four choices of the number of retained factors  $K$ . For  $K=1$ , the eigenvector (black) captures the low-frequency behavior of the column-variance, but ignores the narrow alpha peak and broad beta peak. For  $K=2$ , the first eigenvector (black) behaves similarly, and the second eigenvector (blue) captures the alpha peak.



**Fig. 2.** Scree plots for spectral PCA: (a) PCA-ANOVA without baseline subtraction, (b) PCA-ANOVA with baseline subtraction, (c) STAT-PCA in difference mode, and (d) STAT-PCA in common mode. Filled dots indicate MPL and PA choices for factor retention; in all cases  $MPL < PA$ .

The effects of rotation are minimal in this case. For  $K=3$ , the roles of the first two eigenvectors are unchanged, and the third eigenvector (red) is quite complicated, including peaks at 1 Hz, 4 Hz, 11 Hz, and 29 Hz, with non-uniform signs. Again the effects of rotation are minimal. For  $K=4$ , the roles of the first two eigenvectors are again unchanged. Although the third eigenvector (red) is still complicated, it is simplified slightly by losing much of its peak at 29 Hz after rotation. The fourth eigenvector (green) isolates 29 Hz almost exclusively after rotation. We used these arguments to support the choice of  $K=4$  in this example, and note that this choice is well within the upper limit recommended by PA (Fig. 2c). A similar procedure was used to select  $K$  for each step of sequential PCA.

#### Retention of 60 Hz noise in PCA-ANOVA

In our initial explorations with PCA-ANOVA we passed the time-varying PSD in both conditions to PCA, without baseline subtraction. We kept frequencies up to 100 Hz to see if any high-gamma activity could be found. We found that without baseline subtraction one of the largest components (second spectral factor, fifth spatial factor, third temporal factor) that survived the ANOVA ( $F(1,48)=5.74$ ;  $p=0.0205$ ) reflected 60 Hz noise. The spectral loading had a single, narrow peak at 60 Hz. The spatial loading was peaked near the ground electrode, consistent with theory (Ferree et al., 2001). The time course of the temporal factor was non-descript. We had not anticipated this result, because ANOVA was supposed to isolate differences between conditions, and the difference in 60 Hz noise between conditions should be negligible.

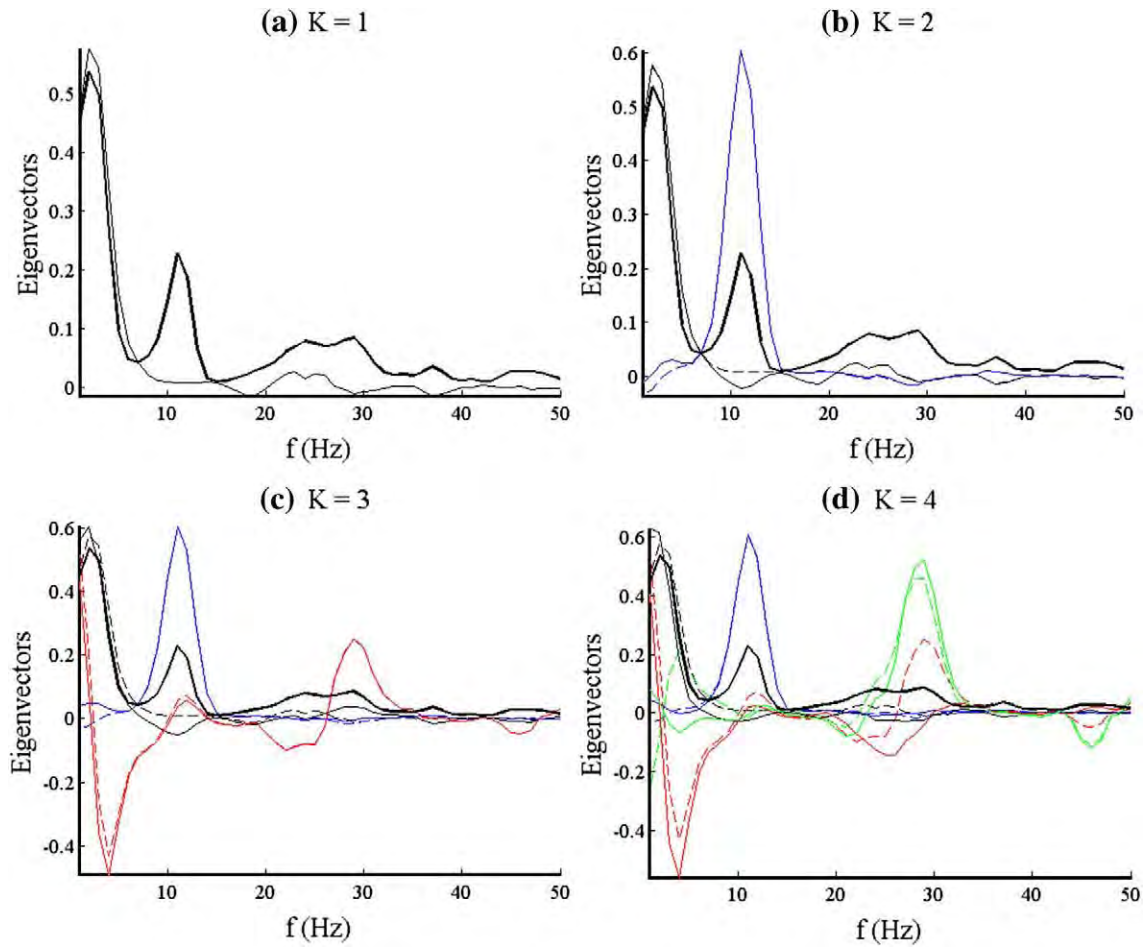
An explanation for why this happens is as follows. When conducting spectral PCA, the data matrix has rows that include

electrodes, time points, conditions (PCA-ANOVA only) and subjects. Prior to forming the covariance matrix, the column-means are subtracted from the data matrix, consistent with the usual definition of covariance matrix. The column-variances are generally non-zero, however, even for the 60 Hz column. Because this feature dominates the covariance matrix, 60 Hz noise emerges as one of the largest factors. While it may be possible to dismiss this, saying that 60 Hz noise is easily filtered, or that one could simply limit the frequency range to 1–50 Hz, we view this as an indication that PCA-ANOVA failed to isolate task-related differences between conditions. Attempting to remedy this problem, we subtracted the baseline power from the moving-window PSD in each condition, in analogy with ERP analysis. We found that the 60 Hz factor no longer reached significance in the ANOVA, and tentatively concluded that we had solved this problem.

#### Sensitivity of PCA-ANOVA to the definition of baseline power

We considered two ways of estimating the baseline power spectrum. The first way was based upon the Welch method, using 50% overlapping windows. In the 1-sec baseline interval, this gives three 0.5-s windows that are nearly independent with the cosine tapering. As shown in Fig. 1, power varies in the baseline interval. The second way was based upon averaging the time-dependent power in all moving windows (0.05 s steps) within the baseline interval. The two ways of computing the baseline power give slightly different results, because the Welch estimate can be seen as three samples among many obtained with sliding windows. Fig. 4 shows scatter plots comparing the two ways of computing the baseline power in a single subject (a) and all subjects (b). In the single-subject case, for this



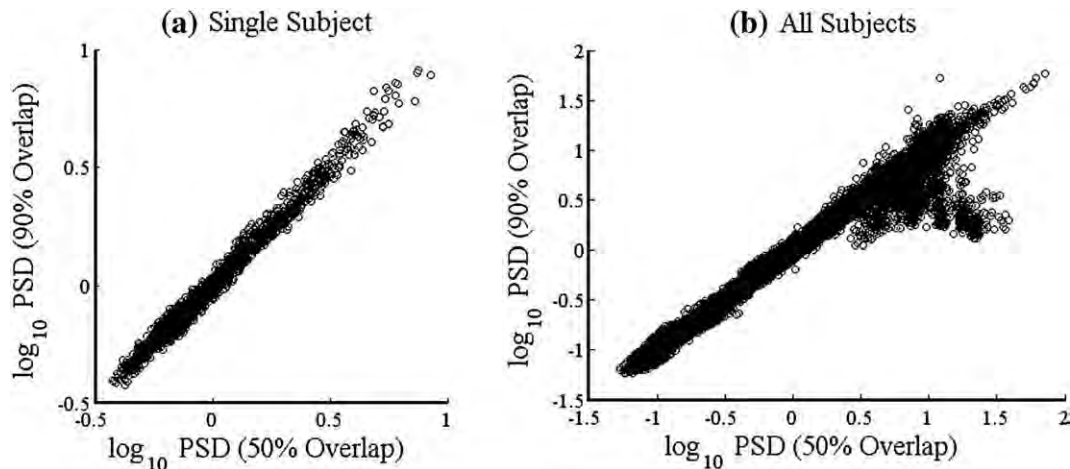


**Fig. 3.** Illustration of strategy for factor retention, shown for spectral STAT-PCA in difference mode. The thick solid curve represents the column-variance of the data array. Thin dashed curves represent eigenvectors before rotation. Thin solid curves represent eigenvectors after rotation. As more factors are retained, their eigenvectors are colored as follows: first (black), second (blue), third (red), fourth (green).

particular subject, the two estimates are highly correlated, suggesting small variability in power during the baseline interval. In the all-subject case, some values deviate far from linearity, revealing large baseline variability in some subjects. Further analysis revealed that these deviants come from one subject mainly, and 2–3 others

secondarily. Yet these subjects are not ‘bad’ per se, as their EEG appear fine, and their responses to the stimulus are visible.

In PCA-ANOVA, the two methods of computing baseline power, Welch 50%-overlap estimate and sliding-window 90%-overlap estimate, gave three factors. Table 1 shows the triplet similarity matrix



**Fig. 4.** Comparison of baseline power in two methods of computing for (a) single subject, (b) all 25 subjects. Horizontal axis shows results for 50% overlap (Welch method). Vertical axis shows results for 90% overlap (averaging all values in moving-window PSD estimate).

Table 1  
PCA-ANOVA: baseline power

90% Overlap	50% Overlap			
		Triplet 1	Triplet 2	Triplet 3
	Triplet 1	0.0004	0.0000	<b>0.0713</b>
	Triplet 2	0.0022	<b>0.0433</b>	0.0137
	Triplet 3	<b>0.0216</b>	0.0234	0.0081

Triplet similarity metrics for two estimates of baseline power in PCA-ANOVA. Both estimates (50% overlap, 90% overlap) gave three significant triplets.

that compares the results. We found  $\max(\Gamma_{ab})=0.0713$ , which implies poor correspondence between the two solutions. We conclude that PCA-ANOVA is highly sensitive to the definition of baseline power.

In STAT-PCA, taking the difference between conditions and keeping only significant differences in single subjects had the effect of removing 60 Hz activity entirely. Even when concatenating across subjects, therefore, STAT-PCA is not affected by 60 Hz. The table for STAT-PCA that would be analogous to Table 1 has ones on the diagonal and zeros elsewhere. We conclude that STAT-PCA is relatively insensitive to the definition of baseline power.

Because the definition of baseline interval, and the method of averaging used to compute baseline power, are rather arbitrary decisions made by the researcher, with little information available to confirm the absolute validity of one choice over another, we argue that any method for analyzing group data should be minimally sensitive to these kinds of choices. The lack of robustness of PCA-ANOVA to the definition of baseline power raises serious concerns about reproducibility of results obtained with this method. In contrast, STAT-PCA isolates task-related spectral changes reliably, which is the stated goal of this entire analysis.

#### Sensitivity of PCA-ANOVA to the deletion of a single subject

When looking for a group effect, it is generally undesirable for one subject to influence the results excessively. In order to assess stability of the factors retained, we calculated PCA-ANOVA for the entire group and compared the results with those obtained by deleting each subject one at a time. When  $N=25$  subjects were used, four factors were deemed significant by ANOVA. When a single subject was removed, three factors were found. That alone raised concern. Furthermore, of the three triplets found when  $N=25$ , only one of those triplets was found when  $N=24$ . We emphasize that the subject removed was *not* one of the subjects that exhibited high baseline variability in Fig. 4b. Indeed, Table 2 was recomputed for each of the 25 subjects separately, and similar results were obtained. To quantify this, we generated the summary statistics as described above and the mean was found to be 0.6870. This implies that PCA-ANOVA is very sensitive

to the deletion of a random subject, making it difficult to generate reproducible results.

#### Robustness of STAT-PCA to the deletion of a single subject

In exact parallel to the procedure used to remove a single subject in PCA-ANOVA, single subjects were removed one at a time from the STAT-PCA analysis and the results compared to the group. In order to make the most direct comparison with PCA-ANOVA, only difference-mode STAT-PCA is shown. Table 3 shows that six triplets were retained for both  $N=25$  and  $N=24$ . Most importantly, the matrix is very nearly the identity matrix, i.e., not only were the same triplets obtained, but they were retained in the same order. Table 3 was recomputed for each of the 25 subjects separately, and similar results were obtained for nearly all subjects. To quantify this, we generated the summary statistics as described above, exactly as they were calculated for PCA-ANOVA. The mean value was 0.9648. The values obtained from PCA-ANOVA and STAT-PCA were compared using a one tailed  $t$ -test and found to be significantly different ( $t(24)=1.8655$ ;  $p=0.0371$ ). This implies that STAT-PCA is highly stable to deletion of a random subject, which helps insure reproducible results.

#### Task-related factors in STAT-PCA

For our group of 25 subjects, STAT-PCA produced two common-mode (CM) triplets and six difference-mode (DM) triplets. Because the goal of this paper is to illustrate the methodology, and establish the internal consistency of the principle components with respect to the original data, only three examples are presented here. A thorough description of all the components and their interpretation as cognitive processes will be presented elsewhere. For each triplet below, thick curves represent factor loadings, and thin curves represent the group-averaged PSD at particular frequencies and electrodes as noted. In topographic plots, the color scale is dark for small values, red for intermediate values, and white for large values. The display of electrodes on the head model is slightly compressed, so that inferior occipital electrodes are also visible.

Fig. 5 shows common-mode triplet CM (1,1,1). The spectral loading has a prominent peak at 11 Hz, with a smaller peak near 2 Hz. The spatial loading is spread over occipital–parietal cortex, and is slightly left-lateralized to peak at electrode PO3. The temporal loading is shown with the group-averaged, common-mode PSD at 11 Hz and electrode PO3 (thin line). Both the loading and data start near zero, fall abruptly until 1 s, then return toward baseline through the rest of the epoch. The temporal loading matches the group-averaged data remarkably well, confirming the internal validity of this factor, and showing that this triplet reflects ERD not ERS.

Fig. 6 shows difference-mode triplet DM (4,1,1). The spectral factor has a single, prominent peak at 29 Hz. The spatial topography has two distinct peaks, bilaterally distributed in frontal electrodes F5 (left) and

Table 2  
PCA-ANOVA: subject deletion

N = 24	N = 25			
		Triplet 1	Triplet 2	Triplet 3
	Triplet 1	0.0000	0.0000	0.0006
	Triplet 2	<b>0.9993</b>	0.0004	0.0001
	Triplet 3	0.0005	<b>0.0934</b>	0.0000
	Triplet 4	0.0001	0.0005	<b>0.0045</b>

Triplet similarity metrics for entire subject group ( $N=25$ ) and one subject deleted ( $N=24$ ) in PCA-ANOVA. Calculations for  $N=24$  gave a different number of significant triplets depending upon the subject that was dropped.

Table 3  
STAT-PCA: subject deletion

N = 24	N = 25					
		Triplet 1	Triplet 2	Triplet 3	Triplet 4	Triplet 5
	Triplet 1	<b>0.9999</b>	0.0052	0.0000	0.0022	0.0074
	Triplet 2	0.0053	<b>0.9999</b>	0.0002	0.0081	0.0022
	Triplet 3	0.0000	0.0002	<b>1.0000</b>	0.0002	0.0003
	Triplet 4	0.0023	0.0079	0.0002	<b>0.9996</b>	0.0170
	Triplet 5	0.0071	0.0023	0.0003	0.0168	<b>0.9993</b>
	Triplet 6	0.0014	0.0000	0.0001	0.0004	0.0016

Triplet similarity metrics for entire subject group ( $N=25$ ) and one subject deleted ( $N=24$ ) in STAT-PCA. Calculations for  $N=24$  gave six significant triplets consistently.

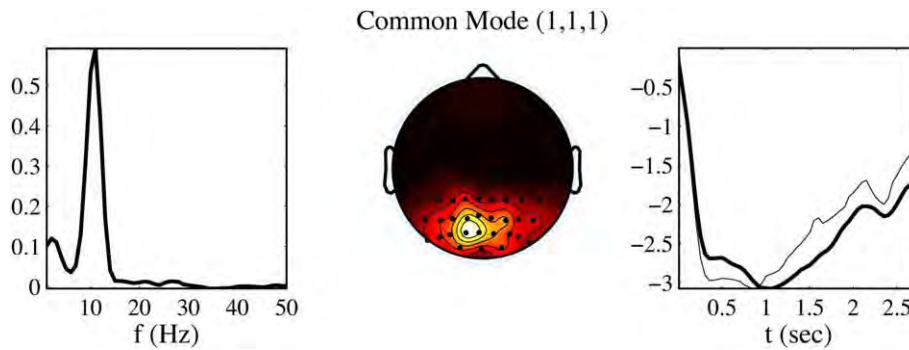


Fig. 5. STAT-PCA triplet CM (1,1,1). The temporal factor (thick line) is plotted with the group-averaged PSD at 11 Hz and electrode PO3 (thin line).

AF4 (right). Because the display of the electrodes is slightly compressed, F5 is actually more lateral than indicated in the graph. The temporal factor (thick solid line) is shown with the group-averaged, difference-mode PSD at 29 Hz, for electrodes F5 (thin solid line) and AF4 (thin dashed line). The time course of this factor and both electrodes are nearly identical. This supports the validity of finding this bifocal map, and suggests functional coupling between these locations.

Fig. 7 shows difference-mode triplets DM (1,1,1) and DM (1,1,2). The spectral factor is peaked at 1 Hz, falling to zero by 9 Hz. This low-frequency power is not merely an artifact due to spectral leakage from the zero-frequency bin, because 1) we used linear detrending to compute power spectra, and 2) difference-mode involves subtracting two conditions and the amount of spectral leakage should not depend strongly upon condition. The topography has a primary peak at electrode PO7, and a secondary peak at electrode AF3, perhaps extending to AF4. The temporal factor (thick solid line) is plotted with the group-averaged, difference-mode PSD at 1 Hz, for electrodes PO7 (thin solid line) and AF3 (thin dashed line). Overall, the temporal behavior of electrode PO7 and electrode AF3 are quite similar, although PO7 is larger for  $t < 0.5$  s. The difference in the temporal factors is that DM (1,1,1) reflects the late behavior that peaks around  $t \approx 1.3$  s, while DM (1,1,2) reflects the early behavior that is confined to  $t < 0.5$  s. An explanation of how and why PCA separated these two temporal factors is given in the Discussion.

## Discussion

Time–frequency analysis of multi-electrode EEG data in cognitive studies yields high-dimensional numerical results spanning space, time, frequency, conditions and subjects. There is a clear need to reduce and summarize these data, with the goal of isolating distinct neural processes involved in the task. Our initial attempt to extend the established technique of PCA-ANOVA to the frequency domain revealed three main shortcomings: 1) isolation of non-task-related

differences between conditions (e.g., 60 Hz noise), 2) sensitivity to the precise definition of baseline power, and 3) sensitivity to the deletion of a single subject. We developed a new approach, called STAT-PCA, which advocates within-subject statistical testing followed by PCA. STAT-PCA was demonstrated to remedy all three of the short-comings of PCA-ANOVA, and yield components that have visible agreement with the group-averaged data. Upon close consideration, it makes intuitive sense that isolating task-related differences in single subjects improves the performance of PCA, because PCA always arranges the largest contributors to variance in its first few components. Only if the interesting data features are also the largest data features will PCA arrange them properly. Furthermore, STAT-PCA requires that a power difference reach significance in the single subject before it can contribute to the covariance matrix, while PCA-ANOVA considers all contributions that may or may not be significant within single subjects. For this reason, we propose that STAT-PCA is conceptually more rigorous than PCA-ANOVA. Finally, STAT-PCA permits the study of activity that is not only different between conditions, but also common to both conditions.

Sequential PCA produces factors that branch like a tree. The first PCA, in this case spectral, gives several factors. For each spectral factor, the second PCA, in this case spatial, gives one or more factors. For each spatial factor, the third PCA, in this case temporal, gives one or more factors. In this way, sequential PCA can tease apart features of the data that differ only at lower levels. An example of this is DM (1,1,1) and DM (1,1,2), which share spectral and spatial factors, but have different temporal factors. At first this separation may seem arbitrary, but it occurred presumably because of structure in the covariance matrix. The fact that this separation is visible in the group-averaged data provides corroborating evidence that this separation is valid. It is therefore a success of STAT-PCA to disambiguate these factors, and we interpret them as distinct physiological processes that likely have distinct interpretations in terms of cognitive processing as well.

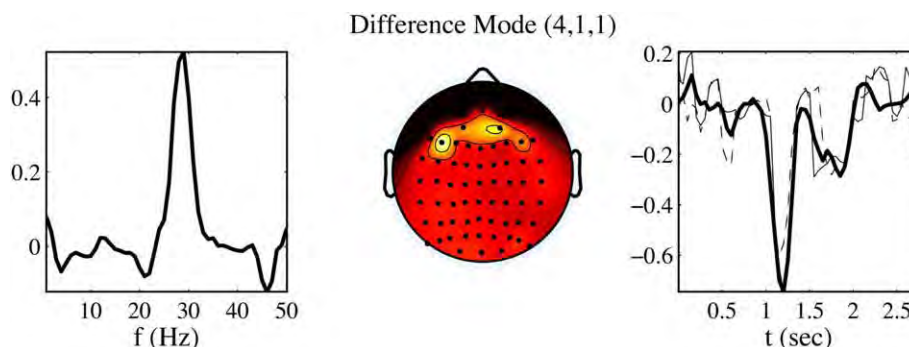
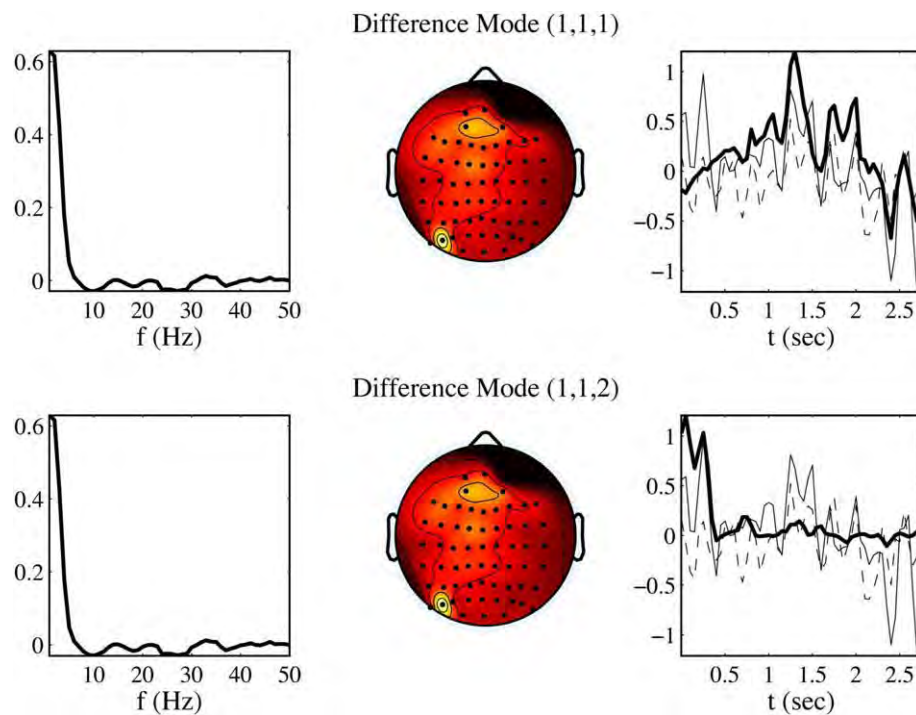


Fig. 6. STAT-PCA triplet DM (4,1,1). The temporal factor (thick solid line) is plotted along with the group-averaged, difference-mode PSD at 29 Hz, for electrodes F5 (thin solid line) and AF4 (thin dashed line).





**Fig. 7.** STAT-PCA triplets DM (1,1,1) and DM (1,1,2). The temporal factor (thick solid line) is plotted with the group-averaged, difference-mode PSD at 1 Hz, for electrodes PO7 (thin solid line) and AF3 (thin dashed line).

Varimax rotation was applied to spectral, spatial, and temporal dimensions. By definition, it produces eigenvectors that are concentrated on a few elements, but those elements need not be proximal to each other. In the spectral domain, most factor loadings had a single, prominent peak, with one or more small sub-peaks. This situation is consistent with prior knowledge that task-related cortical oscillations tend to be relatively narrow-banded, and different bands tend not to overlap. In the spatial domain, Varimax rotation produces maps involving few electrodes, but these electrodes need not be adjacent, as seen in Figs. 6 and 7. In the temporal domain, the loadings tended to be less compact, reflecting sustained neural oscillations during the task. Despite this tendency, temporal factors may be compact, as seen in Fig. 7.

Unlike the standard method PCA-ANOVA, which requires multiple subjects as samples for the ANOVA, a major strength of the new method, STAT-PCA, is the ability to analyze single subjects. Because the goal of the present paper was to compare STAT-PCA with PCA-ANOVA, however, we applied the two methods on equal footing, focusing on group analysis. It might be supposed that, because PCA is a variance-based technique, multi-subject PCA would be more sensitive to inter-subject differences than commonalities. We argue against this possibility on several grounds. First, we showed stability of the factors under deletion of single subjects, thus no single subject (remembering that some subjects can be considered outliers) overly influences the results. Second, in Figs. 5–7 the high correlation between temporal factors (thick lines) and group-averaged power (thin lines) confirms that the STAT-PCA factors reflect the group-averaged behavior. Third, we have begun a follow up study in which we have done single-subject analysis using STAT-PCA, and have found preliminarily that the major factors that emerge for the group are visible in most of the single subjects. Beyond these points, group analysis provides advantages over single-subject analysis, because only in group analysis is the last (temporal) dimension submitted to PCA. As noted above, Fig. 7 shows how, for a particular spectral and spatial factor, the temporal PCA identified two temporal factors. In order to do this last (temporal) PCA, we had to use subjects as

samples. When doing single-subject STAT-PCA, the last (temporal) factors must be obtained as the scores of the previous (spatial) PCA, thus there can be only one temporal factor for each spatial factor. For this reason, it would not have been possible to separate the two temporal factors shown in Fig. 7 in single subjects. Future work will investigate thoroughly the relationship between group and single-subject analyses, especially because the latter is necessary for clinical diagnosis.

A related approach, multi-way or parallel factor (PARAFAC) analysis, has also been applied to reduce space–time–frequency data (Miwakeichi et al., 2004). Because it operates on all three dimensions simultaneously, it avoids the issue of ordering in sequential PCA. Although PARAFAC is receiving much recent attention in the literature, its structure is such that each spectral factor is associated with only one spatial and temporal factor. It seems unlikely that PARAFAC would have separated the two processes shown in Fig. 7. Furthermore, it is often noted that PARAFAC solutions are unique, avoiding the rotation ambiguity in PCA. Practically speaking, however, the use of PARAFAC depends upon several choices, including the number of factors to retain, and constraints between factors in each dimension: orthogonality, positive-definiteness, and compactness. In this sense, the results of PARAFAC analysis are not unique, and more work is required to understand the effects of these constraints and the relationship of PARAFAC with PCA.

Beyond the internal consistencies of STAT-PCA that are the emphasis of this paper, at least one of the triplets found here is consistent with published findings in this task. DM (4,1,1) showed that electrodes F5 and AF3 oscillate in the range 20–35 Hz, and have nearly identical time course. First, this frequency band was found in this same task using intra-cranial electrodes (Slotnick et al., 2002). Second, electrode F5 was found previously by applying PCA-ANOVA to ERP analysis of the same data set (Brier et al., 2008). Third, the time duration (0.75–1.5 s) is the same as that of the ERP. Finding what appears to be the same neural process with ERP and time–frequency analysis gives further credibility to our approach developed here, and provides new ideas for how to clarify its functional role in semantic processing. Because electrodes F5 and AF3 oscillate at the

same frequency and share the same time course, we hypothesize that coherence analysis or phase synchrony analysis (Lachaux et al., 1999) will show these areas to be coupled in the frequency band 20–35 Hz.

In summary, STAT-PCA provides a basis for the reduction of the results of time–frequency analysis of multi-electrode EEG data into concise components that facilitate cognitive interpretation. It represents a paradigm shift for the integration of PCA with statistical testing, by advocating statistical tests in single subjects prior to PCA. In this way, PCA is relegated to a purely descriptive role. As a result of the statistical test occurring first, the factors retained do not need to be subjected to further statistical testing, which previously has been highly subjective. We conclude that STAT-PCA represents an extensible platform for the analysis of event-related spectral changes in cognitive experiments, as well as an adaptive platform for future developments that should include higher-dimensional (i.e., more than two-condition) experimental designs.

## Acknowledgments

The authors thank Cliff Calley and Mary Katherine Reagor for assistance with data acquisition, Priya Xavier and Audrey Chang for assistance with pre-processing, Jeff Spence, Pat Carmack, and Wayne Woodward for discussions about spectral estimation, Gail Tillman, Mandy Maguire, and Joseph Dien for discussions about PCA, and Michael Motes for discussions about stability analysis. This work was supported in part by the Berman Research Initiative, and the Departments of Radiology (Dr. Richard Briggs) and Internal Medicine (Dr. Robert Haley) at the University of Texas Southwestern Medical Center.

## Appendix A

### Implementation of PCA

In each step of the PCA, the time-varying PSD values are arranged in a matrix with variables denoting the PCA dimension, e.g., spectral, spatial, or temporal, and rows denoting the concatenation of the remaining variables, conditions (PCA-ANOVA only), and subjects. Given a data matrix  $X$ , where rows denote samples and columns denote variables, the first step of PCA is to subtract the column-means, i.e., for each column separately the mean across rows is subtracted:

$$Y = X - \bar{X}.$$

(Because rows denote samples, the column-mean may be interpreted as the sample-mean for each variable.)

This mean-subtracted data matrix  $Y$  has a singular value decomposition:

$$Y = USV^T$$

where  $U$  is an orthogonal matrix with columns equal to the left eigenvectors,  $S$  is a diagonal matrix of singular values, and  $V$  is an orthogonal matrix with columns equal to the right eigenvectors. By definition, an orthogonal matrix satisfies  $U^T U = 1$ ; the orthogonality of the matrix  $U$  is accomplished by the orthonormality of each column of  $U$ . The number of non-zero singular values is equal to the lesser of the number of samples or variables.

The covariance matrix is defined:

$$C = Y^T Y = [USV^T]^T [USV^T] = VS^2 V^T$$

where the normalization factor (equal to the number of samples minus one) has been ignored in its definition, because an overall constant does not affect the decomposition. The last equality arises from the orthogonality of the matrix  $U$ . In the language of PCA, the right eigenvectors are called the factor loadings, and the elements of  $S^2$  are called the eigenvalues.

The factor scores  $W$  are obtained by projecting the data onto the orthonormal basis comprised of the factor loadings:

$$W = YV = US$$

where the second equality results from the SVD of  $Y$ . The factor scores are not merely the left eigenvectors, but include the singular values. In this way, the weight, i.e., the singular value, with which each eigenvector in  $U$  enters into the mean-subtracted data is included in the corresponding score vector.

Varimax rotation is applied to the factor loadings  $V$ , resulting in the rotated factor loadings:

$$V' = RV.$$

Varimax rotation preserves inner products between the eigenvectors in  $V$ , i.e., orthogonality and normalization, so  $R$  is an orthogonal matrix. Many applications of PCA consider only the factor loadings, so the effect of rotation on the factor scores is not considered. In the results presented here, the factor scores were derived by projecting the mean-subtracted data onto the rotated factor loadings:

$$W' = YV'.$$

Only in this way can the original data be considered to be comprised of the resulting factor scores and factor loadings:

$$W' V'^T = YV' V'^T = Y.$$

Keeping track of the rotation matrix  $R$  explicitly:

$$\begin{aligned} W' V'^T &= [YRV][RV]^T \\ &= YRVV^T R^T \\ &= YRR^T \\ &= Y \end{aligned}$$

where the third equality arises from the orthogonality of  $V$ , and the last equality arises from the orthogonality of  $R$ . In each step of the PCA, therefore, a subset of the factor loadings were retained and rotated. For each rotated factor loading, the corresponding factor score was computed by projecting the data matrix onto that loading. Each of these 'rotated' scores was reshaped to form a new data matrix for the next step of PCA.

## References

- Assaf, M., Calhoun, V.D., Kuzu, C.H., Kraut, M.A., Rivkin, P.R., Hart, J., Pearlson, G.D., 2006. Neural correlates of the object-recall process in semantic memory. *Psychiatry Research. Neuroimaging* 147, 115–126.
- Bernat, E.M., Williams, W.J., Gehring, W.J., 2005. Decomposing ERP time–frequency energy using PCA. *Clin. Neurophysiol.* 116, 1314–1334.
- Bokil, H., Purpura, K., Schoffelen, J.M., Thomson, D., Mitra, P., 2007. Comparing power spectra and coherences for groups of unequal size. *J. Neurosci. Methods* 159, 337–345.
- Brier, M.R., Maguire, M.J., Tillman, G.D., Hart, J., Kraut, M.A., 2008. Event-related potentials in semantic memory retrieval. *Journal of the Int. Neuropsychol. Soc.* 14 (5), 815–822.
- Cattell, R.B., 1966. The scree test for the number of factors. *Multivariate Behav. Res.* 1, 245–276.
- Curry, S.H., Cooper, R., McCallum, W.C., Pocock, P.V., Papakostopoulos, D., Skidmore, S., Newton, P., 1983. The principle components of auditory target detection. In: Gaillard, A.W.K., Ritter, W. (Eds.), *Tutorials on ERSP Research: Endogenous Components*. North-Holland Publishing Company, Amsterdam, pp. 79–117.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent components analysis. *J. Neurosci. Methods* 134, 9–21.
- Dien, J., 1998. Addressing misallocation of variance in principal components analysis of event-related potentials. *Brain Topogr.* 11 (1), 43–55.
- Dien, J., Frishkoff, G.A., 2005. Principal component analysis of ERP data. In: Handy, T.C. (Ed.), *Event-Related Potentials: A Methods Handbook*. MIT Press.
- Dien, J., Spencer, K.M., Donchin, E., 2003. Localization of the event-related potential novelty response as defined by principal components analysis. *Cogn. Brain Res.* 17, 637–650.
- Dien, J., Beal, D.J., Berg, P., 2005. Optimizing principal components analysis of event-related potentials: matrix type, factor loading weighting, extraction, and rotations. *Clin. Neurophysiol.* 116, 1808–1825.

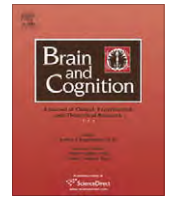


- Dien, J., Khoe, W., Mangun, G.R., 2007. Evaluation of PCA and ICA of simulated ERPs: Promax vs. Infomax rotations. *Hum. Brain Mapp.* 28, 742–763.
- Ding, M., Bressler, S.L., Yang, W., Liang, H., 2000. Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. *Biol. Cybern.* 83, 35–45.
- Donchin, E., 1966. A multivariate approach to the analysis of event-related potentials. *IEEE Trans. Biomed. Eng.* 13 (3), 131–139.
- Donchin, E., Heffley, E., 1979. Multivariate analysis of event-related potential data: a tutorial review. In: Otto, D. (Ed.), *Multidisciplinary Perspectives in Event-related Potential Research*. EPA 600/9-77-043. US Government Printing Office, Washington, DC, pp. 555–572.
- Durka, P.J., Zygierevicz, J., Klekowicz, H., Ginter, J., Blinowska, K.J., 2004. On the statistical significance of event-related EEG desynchronization and synchronization in the time-frequency plane. *IEEE Trans. Biomed. Eng.* 51 (7), 1167–1175.
- Ferree, T.C., 2006. Spherical splines and average referencing in scalp EEG. *Brain Topogr.* 19 (1/2), 43–52.
- Ferree, T.C., Luu, P., Russell, G.S., Tucker, D.M., 2001. Scalp electrode impedance, infection risk, and EEG data quality. *Clin. Neurophys.* 112, 536–544.
- Glorfeld, L.W., 1995. An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educ. Psychol. Meas.* 55, 377–393.
- Hayton, J.C., Allen, D.G., Scarpello, V., 2004. Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. *Organ. Res. Methods* 7, 191–205.
- Horn, J.L., 1965. A rationale and test for the number of factors in factor analysis. *Psychometrika* 32, 179–185.
- Junghofer, M., Elbert, T., Tucker, D.M., Braun, C., 1999. The polar average reference effect: a bias in estimating the head surface integral in EEG recording. *Clin. Neurophysiology* 110 (6), 1149–1155.
- Lachaux, J.P., Rodriguez, E., Marinier, J., Varela, F.J., 1999. Measuring phase synchrony in brain signals. *Hum. Brain Mapp.* 8, 194–208.
- Kalcher, J., Pfurtscheller, G., 1995. Discrimination between phase-locked and non-phase-locked event-related EEG activity. *Electroenceph. and Clin. Neurophysiol.* 94, 381–384.
- Kayser, J., Tenke, C.E., 2003. Optimizing PCA methodology for ERP component identification and measurement: theoretical rationale and empirical evaluation. *Clin. Neurophysiol.* 114, 2307–2325.
- Miwakeichi, F., Martinez-Montes, E., Valdes-Sosa, P.A., Nishiyama, N., Mizuhara, H., Yamaguchi, Y., 2004. Decomposing EEG data into space–time–frequency components using Parallel Factor Analysis. *NeuroImage* 22, 1035–1045.
- Mocks, J., Verleger, R., 1991. Multivariate methods in biosignal analysis: application of principal component analysis to event-related potentials. In: Weitekamp, R. (Ed.), *Digital Biosignal Processing*. Elsevier, Amsterdam, pp. 399–458.
- Nunez, P.L., 1981. *Electric Fields of the Brain*. Oxford University Press.
- Perrin, F., Pernier, J., Bertrand, O., Echallier, J.F., 1989. Spherical splines for scalp potential and current density mapping. *Electroencephalogr. Clin. Neurophysiol.* 72, 184–187.
- Pfurtscheller, G., Lopes da Silva, F.H., 1999. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857.
- Slotnick, S.D., Moo, L.R., Kraut, M.A., Lesser, R.P., Hart, J., 2002. Interactions between thalamic and cortical rhythms during semantic memory recall in human. *Proc. Natl. Acad. Sci.* 99 (9), 6440–6443.
- Spencer, K.M., Dien, J., Donchin, E., 1999. A componential analysis of the ERP elicited by novel events using a dense electrode array. *Psychophysiology* 36, 409–414.
- Spencer, K.M., Dien, J., Donchin, E., 2001. Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology* 38, 343–358.
- Tenke, C.E., Kayser, J., 2005. Reference-free quantification of EEG spectra: combining current source density (CSD) and frequency principal components analysis (fPCA). *Clin. Neurophysiol.* 116, 2826–2846.
- Thomson, D.J., Chave, A.D., 1991. Jackknifed error estimates for spectra, coherences, and transfer functions. In: Haykin, S. (Ed.), *Advances in Spectrum Estimation*. Prentice Hall, pp. 58–113.
- Truccolo, W.A., Ding, M., Knuth, K.H., Nakamura, R., Bressler, S.L., 2002. Trial-to-trial variability of cortical evoked responses: implications for the analysis of functional connectivity. *Clin. Neurophysiol.* 113 (2002), 206–226.
- Turner, N.E., 1998. The effect of common variance and structure pattern on random data eigenvalues: implications for the accuracy of parallel analysis. *Educ. Psychol. Meas.* 58, 541–568.
- Welch, P.D., 1967. The use of fast Fourier transform for estimation of power spectra: a method based on time-averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* 15 (2), 70–73.
- Wood, C.C., McCarthy, G., 1984. Principal component analysis of event-related potentials: simulation studies demonstrate misallocation of variance across components. *Electroenceph. and Clin. Neurophysiol.* 59, 249–260.
- Zhu, M., Ghodsi, A., 2006. Automatic dimensionality selection from the screen plot via the use of profile likelihood. *Comput. Stat. Data Anal.* 51, 918–930.
- Zwack, W.R., Velicer, W.F., 1986. Comparison of five rules for determining the number of components to retain. *Psychol. Bull.* 99 (3), 432–442.



Contents lists available at ScienceDirect

## Brain and Cognition

journal homepage: [www.elsevier.com/locate/b&c](http://www.elsevier.com/locate/b&c)

# The influence of perceptual and semantic categorization on inhibitory processing as measured by the N2–P3 response

Mandy J. Maguire<sup>a,b,\*</sup>, Matthew R. Brier<sup>a</sup>, Patricia S. Moore<sup>a</sup>, Thomas C. Ferree<sup>c</sup>, Dylan Ray<sup>a</sup>, Stewart Mostofsky<sup>d,f,g</sup>, John Hart Jr.<sup>a,e</sup>, Michael A. Kraut<sup>h</sup>

<sup>a</sup> Center for BrainHealth, University of Texas at Dallas, United States

<sup>b</sup> Callier Center for Communication Disorders, University of Texas at Dallas, United States

<sup>c</sup> Department of Radiology, University of Texas Southwestern Medical Center at Dallas, United States

<sup>d</sup> Laboratory of Neurocognitive and Imaging Research, Kennedy Krieger Institute, United States

<sup>e</sup> Department of Neurology, University of Texas Southwestern Medical Center at Dallas, United States

<sup>f</sup> Department of Neurology, Johns Hopkins University School of Medicine, United States

<sup>g</sup> Department of Psychiatry, Johns Hopkins University School of Medicine, United States

<sup>h</sup> Department of Radiology, Johns Hopkins University School of Medicine, United States

## ARTICLE INFO

### Article history:

Accepted 31 August 2009

Available online 20 September 2009

### Keywords:

Event Related Potentials

Inhibition

N2–P3

Go–NoGo

Object–animal categorization

Visual categorization

Conceptual difficulty

## ABSTRACT

In daily activities, humans must attend and respond to a range of important items and inhibit and not respond to unimportant distractions. Our current understanding of these processes is largely based on perceptually simple stimuli. This study investigates the interaction of conceptual-semantic categorization and inhibitory processing using Event Related Potentials (ERPs). Participants completed three Go–NoGo tasks that increased systematically in the degree of conceptual-semantic information necessary to respond correctly (from single items to categories of objects and animals). Findings indicate that the N2 response reflects inhibitory processing but does not change significantly with task difficulty. The P3 NoGo amplitude, on the other hand, is attenuated by task difficulty. Further, the latency of the peak of the P3 NoGo response elicited by the most difficult task is significantly later than are the peaks detected during performance of the other two tasks. Thus, the level of complexity of conceptual-semantic representations influences inhibitory processing in a systematic way. This inhibition paradigm may be a key for investigating inhibitory dysfunction in patient populations.

© 2009 Elsevier Inc. All rights reserved.

## 1. Introduction

Everyday functioning requires the ability to successfully inhibit irrelevant stimuli, thoughts, and behaviors (Logan & Cowan, 1984; Posner & DiGirolamo, 1998). To date, research on response inhibition has started to localize some of the basic neural processes associated with this behavior (Folstein & van Petten, 2008 or Mostofsky & Simmonds, 2008 for review). However, despite the real world implications of successful and unsuccessful inhibition, as reported in Attention Deficit Hyperactivity Disorder (Barkley, 1997; Luu & Tucker, 2001) and healthy aging (Hasher & Quig, 1997), little work has focused on how response inhibition changes as tasks become conceptually more abstract: for example, knowing to stop the car for red lights, small children, or a stray dog, but not for a few leaves blowing across the street. Amongst the most commonly documented manifestations of inhibitory processing are the Event Re-

lated Potentials (ERPs) associated with the Go–NoGo task, in which participants press a button for one type of stimuli and withhold a button response for a second type of stimuli. This task provides a reliable index of inhibitory processes (Perner, Lang, & Kloo, 2002; Simpson & Riggs, 2006; Weintraub, 2000) and elicits predictable changes in the N2 and P3 ERP components. Further, the Go–NoGo task has been used to effectively measure abstract and rapid object categorization (Kincses, Chadaide, Varga, Antal, & Paulus, 2006; Siakaluk, Buchanan, & Westbury, 2003; VanRullen & Thorpe, 2001). The goal of this study is to investigate how inhibitory processing changes as the cognitive demands necessary to respond become systematically less perceptual and more conceptual-semantic in nature.

To date, ERP research has identified two components that relate to inhibitory processing, the N2 and P3 components. Both of these components display larger amplitudes when inhibiting a motor response compared to what is elicited during execution of the response. The N2 is found over fronto-central areas, peaking around 250 ms after stimulus presentation. The P3 is a fronto-central component peaking around 300 ms after stimulus presentation. The

\* Corresponding author. Address: Callier Center for Communication Disorders, University of Texas at Dallas, 1966 Inwood Rd., Dallas, TX 75235, United States. Fax: +1 214 905 3006.

E-mail address: [mandy.maguire@utdallas.edu](mailto:mandy.maguire@utdallas.edu) (M.J. Maguire).

relationship between the N2, the P3, and inhibitory processing is debated (Bruin, Wijers, & van Staveren, 2001; Lavric, Pizzagalli, & Forstmeier, 2004; Smith, Johnstone, & Barry, 2006, 2007, 2008). Some argue that the inhibitory processes are manifested in the N2 (Ciesielski, Harris, & Cofer, 2004; Jodo & Kayama, 1992; Kopp, Mattler, Goertz, & Rist, 1996; van Veen & Carter, 2002), while others argue that the P3 is a more likely candidate as a measure of inhibition (Bruin et al., 2001; Smith et al., 2006, 2007, 2008). There is, however, a general consensus that both are markers of inhibition to some degree (Smith et al., 2007). The present study does not aim to differentiate these components, but to uncover whether and how each of them is influenced by conceptual-semantic processing.

The literature addressing the influence of stimulus type and stimulus presentation on the inhibitory N2 and P3 components has been varied in regard to methodologies and outcomes. Increasing task difficulty by methodological manipulations such as increasing the speed with which participants had to respond (Jodo & Kayama, 1992) resulted in an increase in the N2 NoGo amplitude but no change in the P3. Similarly, Nieuwenhuis, Yeung, and Cohen (2004) found that as the perceptual overlap between the Go and NoGo stimuli increased (for example, discriminating T from F) the amplitude of the N2 NoGo response also increased. When the task difficulty has been more semantic as opposed to perceptual, i.e. stopping on the word “go” and going on the word “stop”, a decrease in the N2 and P3 amplitudes has been reported (Schapkin, Falkenstein, Marks, & Griefahn, 2006). Thus, the limited research to date indicates that while increases in perceptual difficulty may result in increases in the N2 responses, conceptual and semantic increases in task demand appear to result in a decrease in the amplitudes of the N2 and P3 responses.

The goal of this study is to delineate how inhibitory processes are influenced by conceptual-semantic processing in a systematic way. To investigate this question we used object categorization, a core cognitive ability that calls on both perceptual and conceptual-semantic knowledge (French, 1995; French, Mareschal, Mermillod, & Quinn, 2004; Goldstone & Barsalou, 1998; Murphy & Kaplan, 2000; Schyns, Goldstone, & Thibaut, 1998). Categories at the basic level, such as dogs, horses, and birds can be formed and differentiated from one another based on perceptual features (French et al., 2004). In fact, it has been found that pre-lingual infants as young as 3–4 months of age can group dogs as distinct from cats based on quite complex commonalities and differences in their component visual features (French et al., 2004). However, often times we need categories that extend beyond perceptual similarities. One example would be the category ‘animals’, which includes disparate perceptual entities such as a snake and an elephant. In forming these equally important categories humans must rely more heavily on abstract conceptual and semantic processes.

Using ERP responses to the Go–NoGo paradigm for abstract categories of objects and animals, Thorpe and Fize (1996), reported differences between Go (objects) and NoGo (animals) trials by about 150 ms over frontal areas, even using animals in complex scenes compared to a wide range of non-animal target items as the stimuli. The authors concluded that this was a marker of the human capacity to categorize perceptually dissimilar items into conceptually meaningful categories at an “ultra-rapid” pace. Given the use of a Go–NoGo paradigm, it is likely that the differences reported in this study were also related to the N2 inhibitory process, meaning that the two processes of inhibition and categorization operate contemporaneously, and are mediated in the same or closely approximated brain regions. However, the study did not employ multiple levels of categorization from perceptually-based to abstract. As a result, the nature of this interaction remains unclear.

In this study, we address the issue of how inhibitory processes and conceptual-semantic complexity interact by using three different inhibition Go–NoGo tasks which each require different levels of

semantic abstraction to make a correct response. Each includes Go items presented 80% of the time and NoGo items presented 20% of the time. The “Single” task includes one image of a car (Go) and one image of a dog (NoGo). Because the identical images are repeated, the perceptual properties of the items stay identical, limiting the need to categorize across distinct images to respond correctly. The “Multiple” task contains multiple pictures of cars (Go) and multiple pictures of dogs (NoGo) that vary in orientation and subordinate type of item (e.g., SUVs, trucks, convertibles, beagles, great danes, golden retrievers). Thus, correctly responding requires some item-level identification across category exemplars that can be accomplished by focusing on common perceptual features of the items (legs, eyes, wheels, windshields) and then grouping them based on their semantic item-level representation (dog or car). It is important to note that responses to this task can also be made on the supra-ordinate level of object or animal, but only one basic item from each supra-ordinate (i.e., only dogs, as opposed to dogs in addition to other animals) category was used. The “Semantic” task is the least perceptual, most conceptual-semantic task. It includes a wide range of perceptually dissimilar non-animals (Go), from the categories of clothing, tools, furniture, and vehicles and a wide range of animals (NoGo), including a spider, a worm, a lobster, and a dog. Although there are some perceptual features that can be used to identify the animals as distinct from all other items, such as eyes and legs, these were kept to a minimum, biasing toward a focus on the semantic categorization of items. In addition, we have included a more standard Go–NoGo design (“Standard”), using arrows (Go) and octagons (NoGo) which should elicit pre-learned inhibitory responses as a check that the expected inhibitory responses are being elicited.

This design allows us to address our primary question: do inhibitory processes as measured by the N2 and P3 components vary based on the amount of perceptual and conceptual-semantic information necessary to correctly respond? We predict that each of these tasks, regardless of difficulty, will require inhibitory processing that will result in larger amplitudes for the NoGo items compared to Gos in the N2 and P3. By comparing across tasks we can investigate the interaction between inhibitory processes and conceptual-semantic information. We predict that if the differences between Single, Multiple, and Semantic are the result of a difficulty in distinguishing the Go items from NoGo items on a perceptual level, the N2 NoGo amplitudes will decrease systematically with difficulty. Specifically, Semantic would be significantly larger than Multiple which would be significantly larger than Single. In this case, there would be no corresponding changes in the P3. However, we predict that these differences will not be perceptual, but conceptual in nature, which would influence the P3 amplitude, particularly a decrease with semantic-conceptual difficulty as well as a decrease in the N2 amplitude.

## 2. Methods

### 2.1. Subjects

Thirty-five subjects were recruited from the University of Texas at Dallas community via word of mouth and web-based advertising. All subjects were between the ages of 18 and 31. The subjects were all college students with at least 12 years of education. Subjects were screened, per exclusion criteria, to be free from history of traumatic brain injury and other significant neurological issues (CVA, seizure disorders, history of high fevers, tumors, or learning disabilities). Exclusion criteria also included left-handedness, use of alcohol or other controlled substances within 24 h of EEG administration, and medications other than over-the-counter analgesics and birth-control pills.

Subjects were not included in the final analysed data set if one or more tasks included major movement artifact, failure to comprehend the instructions, errors of omission or commission above 20%, or excess noise ( $N = 3$ ). Thus, data from 32 subjects across all four tasks were fully analysed. The pool of included subjects was 14 males and 18 females. Informed written consent was collected from each subject according to the rules of the Institutional Review Board of The University of Texas at Dallas. This study was conducted according to the Good Clinical Practice Guidelines, the Declaration of Helsinki, and the US Code of Federal Regulations.

## 2.2. Stimuli

Four different tasks were developed, each following the basic Go–NoGo paradigm. In each task, there were 160 (80%) ‘Go’ stimuli for which the subject was to press a button and 40 (20%) ‘NoGo’ stimuli for which the subject was instructed to withhold a response. In each of the four tasks stimuli were presented for 300 ms followed by a fixation point (+) for 1700 ms. All of the stimuli were black line drawings fitted to a white  $600 \times 600$  pixel square.

As can be seen in Fig. 1, in the first task (Standard) the Go stimulus was an arrow and the NoGo stimulus was an octagon. It was predicted that these images would include semantic information about the required response, with an arrow to press a button and an octagon (like a stop sign) to not press the button. Thus, the items were predicted to be cognitively rudimentary compared to the stimuli used in the subsequent tasks. In the second task (Single), the Go stimulus was a line drawing of a car and the NoGo stimulus was a line drawing of a dog. Because the exact same images were repeated multiple times, no categorization of perceptually variable items was necessary for making a response. As such, this is the most perceptually-based condition. In the third task (Multiple), the Go stimuli included 40 cars and the NoGo stimuli included 10 dogs, which were each repeated four times each. To increase the amount of perceptual similarity, all of the images of cars and dogs were drawn as side-views. The images of cars included at least two wheels, front seat and backseat windows and side-view mirrors. The images of dogs included at least two legs, an eye, ears, a nose and a tail. Unlike the Single condition the size and orientation of the items differed across stimuli, but each exemplar of the

dog and car categories includes these perceptual cues. Each subject saw each of the cars and the dogs presented in a random order.

In the final task (Semantic), the Go stimuli consisted of 160 objects (40 food items, 40 cars, 20 clothing items, 20 kitchen items, 20 body parts, and 20 tools) and the NoGo stimuli consisted of 40 animals of varying visual typicality. These ranged from animals seen frequently in daily life, such as dogs and cats, to those seen less frequently such as lobsters and elephants. To decrease the reliance on perceptual cues when performing this task, features often seen in animals were decreased in presentation frequency in the animal condition and increased in the object condition. For example, some of the animals (worm, butterfly, fish) had no legs, while some of the objects (chair, table, a pair of pants) did have legs. Similarly multiple animals without discernable eyes were included, such as the worm, the fly and the lobster. By limiting the perceptual cues that are available in the Multiple and Single conditions participants were required to rely more on semantic-conceptual information to make category responses.

The images were from the standardized picture sets Snodgrass and Vanderwart (1980) and the Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1983) or were created in-house. All attempts were made to keep the images in a similar style and the line thickness of all pictures was adapted to range between 5 pixels and 14 pixels to maintain similarity in visual complexity.













Six randomizations were created to eliminate order effect within each task and the order in which the subjects performed the tasks was counterbalanced to mitigate order effects. The instructions were identical across tasks with only the important object names changed, i.e., “You are going to see pictures of cars (objects) and dogs (animals). You are to press the button for all cars (objects) but do not press the button for anything else”. Instructions were given verbally and then displayed on the computer screen prior to each task.

## 2.3. Data acquisition and preprocessing

Continuous EEG was recorded from 64 silver/silver-chloride electrodes mounted within an elastic cap (Neuroscan Quickcap) which are placed according to the International 10–20 electrode placement standard (Compumedics, Inc.). The data was collected using a Neuroscan SynAmps2 amplifier and Scan 4.3.2 software sampled at 1 kHz with impedances typically below 10 k $\Omega$ . Blinks and eye movement were monitored via two electrodes, one mounted above the left eyebrow and one mounted below the left eye. The data were processed to remove ocular and muscle artifacts in the following way. First, poorly functioning electrodes were identified visually and removed. Second, eye blink artifacts were removed by a spatial filtering algorithm in the Neuroscan Edit software using the option to preserve the background EEG. Third, time segments containing significant muscle artifacts or eye movements were rejected. The continuous EEG data were band-pass filtered from 0.15 to 30 Hz for analysis. The EEG data were segmented off-line into 2 s epochs spanning 100 ms before to 1500 ms after the presentation of the visual stimuli.

## 2.4. Reference correction

The data were recorded with the ground at AFz and the reference electrode located near the vertex, resulting in small amplitudes over the top of the head. In order to eliminate this effect, the data were re-referenced to the average potential over the entire head, which approximates the voltages relative to infinity. In order to minimize a small bias in the electrode-based average reference (Junghöfer, Elbert, Tucker, & Braun, 1999), a spline-based estimate of the average scalp potential (Ferree, 2006) was computed using spherical splines (Perrin, Pernier, Bertrand, & Echallier,

	Go		No Go	
Standard (repeated)				
Single (repeated)				
Multiple (examples)				
Semantic (examples)				

**Fig. 1.** Sample of stimuli used across all four inhibition tasks. Go items were shown 160 times, or 80%, while NoGo were shown 40% or 20%. For the Standard and Single tasks these specific items were repeated 160 and 40 times. For the Multiple and Semantic conditions there are example items taken from the larger set.



1989). Placing the electrode cap on a realistic phantom head, the electrode coordinates were digitized (Polhemus, Inc.), and these coordinates were used to fit the splines for each subject. In subjects with a small number of bad electrodes, the splines were used to interpolate those electrodes, to yield a total of 62 data channels in every subject.

### 2.5. ERP calculation

For each trial and electrode, the mean amplitude of the prestimulus interval (–100 ms to 0 ms) was subtracted from each time point and those data were averaged across trials to create the ERP. Only those items to which correct responses were given were included in the analysis. The waveforms for a midline frontal electrode (Fz) can be seen in Figs. 2 and 3.

### 2.6. Peak analysis

Fig. 4 depicts the grand mean average across the scalp for Multiple, which is typical of all of the studies. Of note is that the N2 and P3 effects are distinctly visible. The N2 is focused over frontal midline and bilateral sites. The P3, on the other hand, is maximal over a wide fronto-central area.

The ERP topographies for N2 and P3 responses separately were averaged across sets of electrodes displaying maximal responses. Based on previous literature and visual inspection of individual and group data across the four tasks, we defined two time areas of interest. For the N2 component we calculated the peak amplitudes over frontal areas by identifying the peak amplitude between 150 and 300 ms for midline frontal electrodes (FPz, FP1, FP2, AF3,

AF4, Fz, F1, F2) for each condition for each participant then averaging across participants. For the P3 effect, the mean peak voltage at central sites (FCz, FC1, FC2, Cz, C1, C2, CPz, CP1, CP2) was calculated between 300 and 600 ms.

## 3. Results

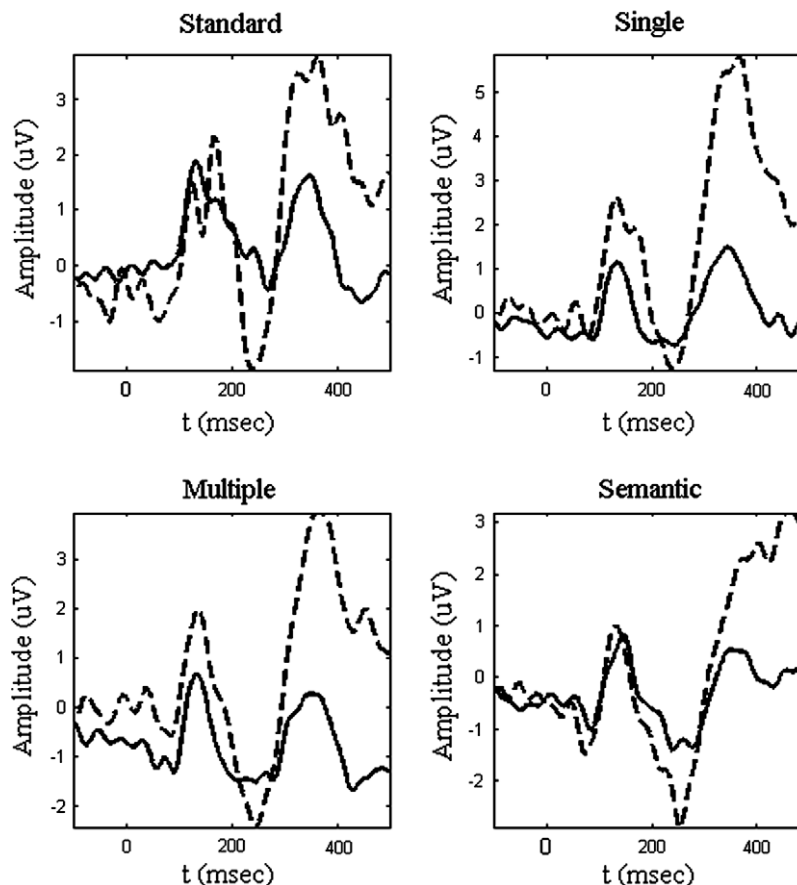
### 3.1. Behavioral

#### 3.1.1. Error rates: errors of omission

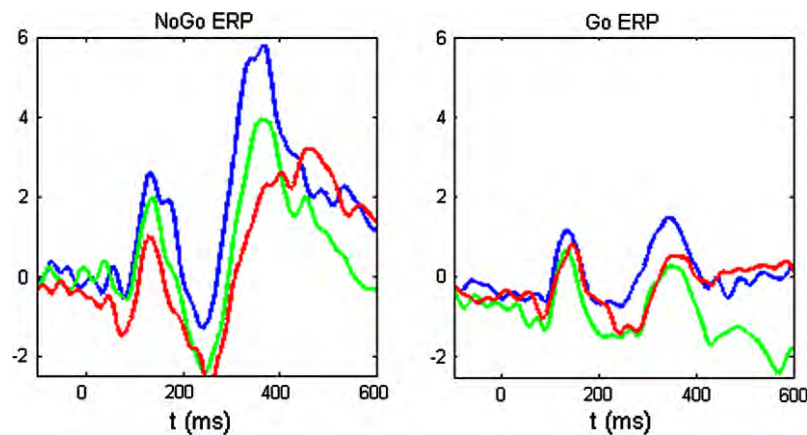
An error of omission is a failure to respond to a Go item. To determine this tendency across tasks, a 4 (task: Standard, Single, Multiple, Semantic)  $\times$  2 (condition: Go, NoGo) repeated-measures ANOVA on the percentage of correct Go responses was calculated. A lower percentage represents more errors of omission. This analysis revealed a significant task difference,  $F(3, 93) = 3.88$ ,  $p = 0.012$ . Further there was a significant linear trend,  $F(1, 30) = 6.82$ ,  $p = 0.014$  in which percent correct decreased with task difficulty, as is depicted to Table 1. Following this trend, directed pair-wise comparisons were performed between each task and the next most difficult (Standard to Single, Single to Multiple, Multiple to Semantic). On a pair-wise comparison level, none of the comparisons were statistically significant after Bonferroni corrections were applied.

#### 3.1.2. Error rates: errors of commission

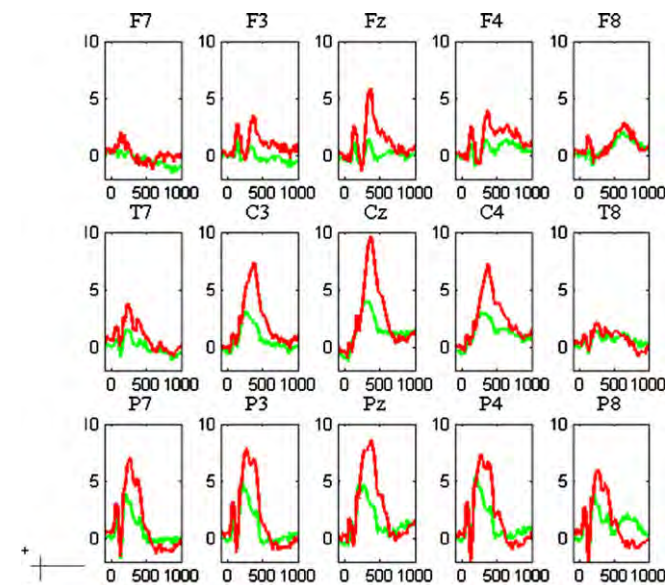
An error of commission is responding with a button press to a NoGo item. In this case a higher percentage indicates fewer errors. To calculate this tendency across tasks a 4 (task: Standard, Single,



**Fig. 2.** N2 and P3 Go and NoGo responses at Fz for Go (solid) and NoGo (dashed) conditions across all tasks. The traditional N2 and P3 components, which display a larger amplitude for NoGo than Go responses, can be seen for each condition.



**Fig. 3.** Grand average at Fz across the three inhibition tasks: Single (blue), Multiple (green), and Semantic (red). As can be seen in this depiction there is a decrease in the amplitude of the P3 with task difficulty. A similar trend in the N2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 4.** Grand average of selected electrodes of interest for the Multiple Task: Go (green) NoGo (red). As can be seen in the figure, the inhibitory Go–NoGo effect was largest over the frontal areas. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 1**  
Means and standard deviations of behavioral results across tasks.

Experiment	Go correct (%)	NoGo correct (%)	Reaction times (ms)
Standard	97.5 (4.3)	82.7 (13.4)	307.7 (9.9)
Single	93.9 (17.6)	82.7 (18.9)	293.7 (5.9)
Multiple	93.1 (17.9)	83.3 (18.6)	307.7 (5.8)
Semantic	90.8 (18.2)	82.5 (18.5)	361.1 (10.7)

Multiple, Semantic)  $\times$  2 (condition: Go, NoGo) repeated-measures ANOVA on the percent of correct NoGos revealed no significant differences across tasks. These means are also listed in Table 1.

### 3.1.3. Reaction times

The reaction times for the Semantic task were longer than the others by approximately 50 ms. To investigate the significance of processing differences between the tasks we performed a one-way ANOVA of the reaction times for Go items. The ANOVA re-

vealed a main effect of task,  $F(3, 93) = 11.32$ ,  $p < 0.0001$ . Subsequent paired samples  $t$ -tests with Bonferroni corrections revealed significant differences between Semantic and each of the other tasks, Standard,  $t(31) = 3.61$ ,  $p = 0.001$ , Single,  $t(31) = 4.74$ ,  $p < 0.001$  and Multiple,  $t(31) = 3.68$ ,  $p < 0.001$ . Other comparisons did not reach significance.

### 3.1.4. Event Related Potentials

A comparison of Go and NoGo waveforms for each task can be seen in Fig. 2, with a comparison across tasks for Go and NoGo responses in Fig. 3. In creating the ERPs only correct responses were included and participants had to have at least 20 artifact-free and correct response epochs per condition. Across tasks there were no significant differences in the number of trials included based on a  $2 \times 3$  repeated-measures ANOVA,  $p > 0.50$ . The mean number of epochs included in the participants' average ERP per condition is as follows, Goes: Single 120.65, Multiple 130.09, Semantic 122.21, NoGoes: Single 28.94, Multiple 30.09, Semantic 30.28. Thus, although there is large variability in the error rates of participants, this did not differentially influence the number of items included in the analysis of each study.

**N2.** The first question was whether the Go–NoGo N2 inhibitory effects elicited by each of these tasks were similar across tasks. The Standard task was assumed to be a baseline against which to compare responses recorded during performance of the three tasks of interest. To determine this for each task the N2 Go and NoGo amplitudes were compared at mid-frontal sites between 150 and 300 ms. As can be seen in Table 2 all of the tasks of interest elicited significantly higher NoGo N2 effects compared to Go effects, in a pattern similar to the Standard task. The only exception is the Semantic task, in which the effect is significant at  $p = 0.01$  without corrections for multiple comparisons, but is not quite significant when these corrections are taken into account.

The next question was how the N2 components were influenced by semantic level of representation. Because the Standard task was

**Table 2**  
N2 amplitude findings across experiments.

Experiment	Go Amp	No Go Amp	$t$ -Stat	$p$ -Value
Standard	−3.13	−5.40	3.68	<0.0001*
Single	−3.34	−4.62	3.00	0.0022*
Multiple	−3.57	−4.79	2.77	0.0005*
Semantic	−3.13	−4.18	2.45	0.01

\*  $p < \alpha$  (0.006) with corrections for multiple comparisons.

a control that is conceptually dissimilar to the three semantic inhibition tasks, in that abstract symbols with preconceived meanings were used instead of objects and animals, it was removed from these comparative analyses. A 3 (task: Single, Multiple, Semantic)  $\times$  2 (condition: Go, NoGo) ANOVA performed on the midline frontal electrodes for the average peak amplitude for the time points between 150 and 300 ms revealed a significant interaction,  $F(2, 62) = 17.78$ ,  $p = 0.04$ . There was also a main effect of task,  $F(2, 62) = 17.08$ ,  $p = 0.008$ , and the main effect of condition nearly reached significance,  $F(2, 62) = 3.77$ ,  $p = 0.06$ . To assess the nature of this interaction two ANOVAs were performed comparing across tasks within each condition, Go and NoGo, but neither ANOVA revealed statistically significant interactions.

**P3.** As with the N2 effects, we first tested the hypothesis that the expected Go–NoGo inhibition effect was elicited by each of our tasks. To do so, *t*-tests were performed on the mean peak amplitudes of the Go conditions compared to the NoGo condition for each condition between 300 and 600 ms over fronto-central sites. As can be seen in Table 3, all of the NoGos elicited significantly larger P3 amplitudes than the Gos, all  $p$ 's  $< 0.0001$ .

Once the expected inhibition effect was verified in comparison to the Standard task, the Standard task was removed from further analyses. To determine how task influenced inhibitory processing as measured by the P3 Go–NoGo a 3 (task: Single, Multiple, Semantic)  $\times$  2 (condition: Go, NoGo) ANOVA was performed. The ANOVA revealed an interaction,  $F(2, 62) = 21.30$ ,  $p < 0.0001$ , as well as main effects of task and condition, all  $p$ 's  $< 0.002$ . To follow up on the interaction two ANOVAs were performed comparing across tasks within each condition. The one-way ANOVA on Go amplitudes revealed no significant differences between tasks. However there were significant differences in the NoGo condition,  $F(2, 62) = 5.77$ ,  $p = 0.005$ . As can be seen in Table 4, subsequent *t*-tests revealed significant differences between Single and Multiple, and Single and Semantic, but not between Multiple and Semantic.

Because the P3 Go effect is often reported to be maximal at central compared to fronto-central areas and this was confirmed based on visual inspection of our grand averages across the tasks, we also performed this analysis on the 300–600 ms epoch over centroparietal electrodes. To determine how task influenced inhibitory processing as measured by the P3 Go–NoGo, a 3 (task: Single, Multiple, Semantic)  $\times$  2 (condition: Go, NoGo) ANOVA was performed at midline centro-parietal sites from 300 to 600 ms after stimulus presentation, which revealed no significant interaction.

Visual inspection of the data also seemed to show that there were large latency differences between tasks for the NoGo conditions (see Fig. 3). As a result, we also performed a one-way between task ANOVA for the NoGo peak latencies between 300 and 600 ms over the fronto-central regions, which revealed significant task differences,  $F(2, 31) = 11.51$ ,  $p = 0.0005$ . As can be seen in Table 5, subsequent *t*-tests revealed that there were significant differences between Single and Semantic, and Multiple and Semantic, but not Single and Multiple. The latency for the Semantic NoGo P3 ( $M = 449.42$  ms,  $SD = 68.07$ ) was significantly longer than for Single ( $M = 404.30$  ms,  $SD = 54.00$ ) or Multiple ( $M = 403.33$  ms,  $SD = 48.91$ ).

**Table 3**  
P3 amplitude findings across experiments.

Experiment	Go Amp	No Go Amp	<i>t</i> -Stat	<i>p</i> -Value
Standard	2.93	6.04	7.21	$<0.0001^*$
Single	3.04	7.35	5.60	$<0.0001^*$
Multiple	1.89	5.75	5.71	$<0.0001^*$
Semantic	2.26	5.04	5.07	$<0.0001^*$

\*  $p < \alpha$  (0.006) with corrections for multiple comparisons.

**Table 4**  
NoGo P3 amplitude differences between tasks.

Comparison	df	<i>t</i> -Value	<i>p</i> -Value
Single – Multiple	31	2.40	0.001*
Single – Semantic	31	3.29	0.001*
Multiple – Semantic	31	0.99	0.16

\*  $p < \alpha$  (0.012) with corrections for multiple comparisons.

**Table 5**  
NoGo P3 latency differences between tasks.

Comparison	df	<i>t</i> -Value	<i>p</i> -Value (one-tailed)
Single – Multiple	31	0.08	0.94
Single – Semantic	31	2.99	0.002*
Multiple – Semantic	31	3.94	0.0002*

\*  $p < \alpha$  (0.012) with corrections for multiple comparisons.

#### 4. Discussion

The focus of this study was to investigate how increased levels of conceptual-semantic processing influence inhibitory processing. We investigated how behavioral measures and known neural markers of inhibitory processing (the N2 and the P3 ERPs) differed across three Go–NoGo tasks that varied systematically based on conceptual-semantic processing necessary to make the inhibitory response. Overall, we found that all three paradigms elicited inhibitory responses similar to our control task of arrows and octagons, but did so in ways that increase our understanding of how inhibitory processes and conceptual-semantic processing interact. Specifically, both the N2 and P3 decreased with task difficulty, though only the P3 does so significantly. This trend suggests the influence of an increase in semantic difficulty as opposed to perceptual difficulty (Schapkin et al., 2003). Had perceptual differences been driving our effects we would have expected an increase in N2 amplitude and no change in P3, as Nieuwenhuis et al. (2004) reported after manipulating stimulus discrimination in a Go–NoGo task.

The known inhibition-related phenomena associated with the N2, i.e., larger amplitude for NoGo compared to Go responses, occurred in all of our semantic inhibition tasks. This is essential to addressing our question of how inhibition and categorization interacted. If there were no N2 NoGo differences, then we would have to conclude that task difficulty slowed the process to such a degree that stopping a response no longer required the invocation of active inhibitory processes. Thus, all of our tasks elicited an inhibitory N2 effect regardless of difficulty, which supports previous literature.

Although there was a trend towards a decrease in the N2 amplitude with task difficulty, post hoc analyses did not reveal any statistically significant differences. In general, however, the trend was as expected.

The amplitude of the P3 to the NoGo stimuli decreased systematically, from Single to Multiple to Semantic, as the conceptual-semantic information needed to make a response became more complex. There are several findings of interest here. First, it was the NoGos and not the Gos that were statistically different across tasks, indicating that inhibitory processes more than attentional ones are influenced. The second point of interest is that the NoGo amplitude decreases, as opposed to increasing, with difficulty. This pattern was also reported by Schapkin et al. (2006). Their task included pressing the button for the word “go” and withholding for “stop” compared to the more difficult, Stroop-like condition in which “go” and “stop” were reversed. The task difference under investigation in the current study is also a conceptual-semantic



one, more so than changes in speed (Jodo & Kayama, 1992) or perceptual similarity between stimuli (Nieuwenhuis et al., 2004). Thus, the fact that the P3 decreases in amplitude, as opposed to becoming larger with task difficulty, is consistent with the idea that the increased difficulty was conceptual as opposed to perceptual.

There are other possible explanations for this effect. The first is that decreased P3 across tasks is caused by different levels of expectation or response preparation across the tasks. Response preparation can result in the changes in the P3 amplitude independently (Smith et al., 2007; Stadler, Klimesch, Pouthas, & Ragot, 2006) and because the P3 may overlap with the CNV or Contingent Negative Variation (Kok, 1988; Simson, Vaughan, & Ritter, 1977). Either cause would result in the same outcome. The higher the predictability of the coming response, the higher the P3 amplitude will be (Bruin et al., 2001; Ruchkin, Sutton, & Tueting, 1975; Stadler et al., 2006). Although the probability of target to non-target items is identical in all three tasks, these effects may be perceptually driven such that as the visual predictability increases, so does the amplitude of the P3. In this case, the Single task has only two stimuli repeated multiple times resulting in a high level of expectancy about any given trial. The Semantic task, on the other hand, has 200 different images, making the predictability of a given trial very low. The Multiple task falls between these two because each stimulus is repeated four times and the items within each category necessarily share important perceptual features. Thus, the decrease in P3 amplitude across these three tasks could be driven by decreased amounts of perceptual similarity within Go and NoGo conditions.

To some degree this argument is similar to our own. We agree that the decrease in P3 amplitude is driven by a decrease in the amount of perceptual information available to form a response. However we argue that this decrease in perceptual cues results in an increase in semantic-conceptual processing which in turn elicits smaller P3 amplitudes. The current data cannot distinguish between increased semantic-conceptual processing and response preparation. The only evidence that we have of an influence beyond stimulus predictability is that in the current study the N2 also displays a slight decrease in amplitude with task difficulty. Response preparation often only influences the amplitude of the P3 with no corresponding N2 difference (Bruin et al., 2001; Smith et al., 2007). Future studies may be able to tease these factors apart by varying item predictability with cues before the stimulus onset, or varying the inter-stimulus interval.

The second alternative explanation for the decrease in P3 amplitude across tasks is that the increase in reaction time, reaction time variability, or peak latency increased variability in the individual P3 peak amplitudes, resulting in a smearing effect and thus a smaller peak amplitude. However, there are reasons to discount this possibility. For one, the Single and Multiple tasks did not differ significantly in reaction times, but the corresponding P3 amplitudes do differ. Secondly, studies focusing on the factors influencing the P3 effect have shown that, based on single trial latency adjustment procedures, reaction time and corresponding P3 latency differences are less influential in determining the P3 amplitude than is task difficulty (see Kok, 2001 for review). As a result, it seems that the systematic decrease in P3 NoGo amplitude from Single to Multiple to Semantic is the effect of increased conceptual-semantic complexity across the tasks.

The relative decreases in the N2 and P3 are especially interesting given the potential effects of repetition. The Single task repeats the identical stimuli 160 and 40 times respectively. In the Multiple task the stimuli are each repeated four times and the Semantic task has no repetitions. If repetition alone were driving the N2 and P3 effects we would actually see the opposite trend, in which with repetition the N2 and P3 amplitudes decrease (Courchesne, Courchesne, & Hillyard, 1978; Polich, 1989; Ravden & Polich, 1998).

The other interesting finding with regard to the P3 was the difference in NoGo peak latency. The latencies were almost identical for Single and Multiple tasks, with a significant increase in the Semantic task. Interestingly, this NoGo latency difference was mirrored in the reaction times for Go responses, in which the Semantic task was significantly longer than the two others. Other studies have reported a difference in P3 latency with stimulus complexity (McCarthy & Donchin, 1981).

Overall, the findings indicate that conceptual-semantic processing does influence inhibitory responses in quite specific and predictable ways. In particular, it appears that inhibition tasks engaging categorization at a superordinate level influence the P3 effect differently than item/object level inhibitory tasks. We remain undecided in the ongoing N2 and P3 inhibitory debate. The fact that N2 is not significantly affected but that P3 is influenced means that by about 300 ms, higher level semantic processing is influencing the inhibitory responses on a neural level. These findings are also of importance because there is a growing body of research focused on applying inhibitory tasks to clinical populations in order to discern neural markers of inhibitory dysfunction. Thus, it is vital that the interactions between conceptual-semantic processing and inhibition be taken into consideration, or systematically tested, when designing studies and interpreting inhibitory processes across populations. Uncovering the range of influence of inhibitory processing deficits in disorders such as Attention Deficit Hyperactivity Disorder, dyslexia, autism, and degenerative diseases will likely necessitate tasks that require higher level cognitive processing to make inhibitory responses. This paper presents a broader range of inhibitory tasks than have typically been used in studies of populations with proposed inhibitory deficits. In addition, using tasks that are more complex than a simple perceptual decision has the potential to address the real world implications of inhibitory deficits evident in disorders such as ADHD.

## Acknowledgments

The authors would like to thank Joshua White, Sandra Chapman, and Jacque Gamino for their comments and help. The funding for this project was provided by a grant from the Sparrow Foundation for the CAARTE program and NIH (NINDS) K02NS044850 and R01NS047781.

## References

- Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin*, 121, 65–94.
- Bruin, K. J., Wijers, A. A., & van Staveren, A. S. J. (2001). Response priming in a go/nogo task: Do we have to explain the go/nogo N2 in terms of response activation instead of inhibition? *Clinical Neurophysiology*, 112(9), 1660–1671.
- Ciesielski, K. T., Harris, R. J., & Cofer, L. F. (2004). Posterior brain ERP patterns related to the go/no-go task in children. *Psychophysiology*, 41, 882–892.
- Courchesne, E., Courchesne, R. Y., & Hillyard, S. A. (1978). The effects of stimulus deviation on P3 waves to easily recognized stimuli. *Neuropsychologia*, 16, 189–199.
- Ferree, T. C. (2006). Spherical splines and average referencing in scalp EEG. *Brain Topography*, 19(1/2), 43–52.
- French, R. (1995). *The subtlety of sameness: A theory and computer model of analogy-making*. Cambridge, MA: MIT Press.
- French, R., Mareschal, D., Mermillod, M., & Quinn, P. C. (2004). The role of bottom-up processing in perceptual categorization by 3- to 4-month-old infants: Simulations and data. *Journal of Experimental Psychology: General*, 133(3), 382–397.
- Folstein, J. R., & van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, 45, 152–170.
- Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, 65, 231–262.
- Hasher, L., & Quig, M. B. (1997). Inhibitory control over no-longer-relevant information: Adult age differences. *Memory and Cognition*, 25, 286–295.
- Jodo, E., & Kayama, Y. (1992). Relation of negative ERP component to response inhibition in a Go/No-go task. *Electroencephalography and Clinical Neurophysiology*, 82, 477–482.



- Junghöfer, M., Elbert, T., Tucker, D. M., & Braun, C. (1999). The polar average reference effect a bias in estimating the head surface integral in EEG recording. *Clinical Neurophysiology*, 110(6), 1149–1155.
- Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *The boston naming test* (2nd ed.). Philadelphia: Lea & Febiger.
- Kincses, T. Z., Chadaide, Z., Varga, E. T., Antal, A., & Paulus, W. (2006). Task-related temporal and topographical changes of cortical activity during ultra-rapid visual categorization. *Brain Research*, 1112, 191–200.
- Kok, A. (1988). Overlap between P300 and Movement-related potentials: A response to Verleger. *Biological Psychology*, 27, 51–58.
- Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, 38, 557–577.
- Kopp, B., Mattler, U., Goertz, R., & Rist, F. (1996). N2, P3, and the lateralized readiness potential in a no-go task involving selective response priming. *Electroencephalography and Clinical Neurophysiology*, 99, 19–27.
- Lavric, A., Pizzagalli, D. A., & Forstmeier, S. (2004). When 'go' and 'nogo' are equally frequent: ERP components and cortical tomography. *European Journal of Neuroscience*, 20(9), 2483–2488.
- Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, 91, 295–327.
- Luu, P., & Tucker, D. M. (2001). Regulating action: Alternating activation of midline frontal and motor cortical networks. *Clinical Neurophysiology*, 112, 1295–1306.
- McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science*, 211, 77–80.
- Mostofsky, S. H., & Simmonds, D. J. (2008). Response inhibition and response selection: Two sides of the same coin. *Journal of Cognitive Neuroscience*, 20, 751–761.
- Murphy, G. L., & Kaplan, A. S. (2000). Feature distribution and background knowledge in category learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 53A, 962–982.
- Nieuwenhuis, S., Yeung, N., & Cohen, J. D. (2004). Stimulus modality, perceptual overlap and the go/no-go N2. *Psychophysiology*, 41, 157–160.
- Perner, J., Lang, B., & Kloo, D. (2002). Theory of mind and self-control: More than a common problem of inhibition. *Child Development*, 73, 752–767.
- Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, 72, 184–187.
- Posner, M. I., & DiGirolamo, G. J. (1998). Executive attention: Conflict, target detection and cognitive control. In R. Parasuraman (Ed.), *The attentive brain* (pp. 401–424). Cambridge, MA: MIT Press.
- Polich, J. (1989). Habituation of P300 from auditory stimuli. *Psychobiology*, 17, 19–28.
- Ravden, D., & Polich, J. (1998). Habituation of P300 from visual stimuli. *International Journal of Psychophysiology*, 30, 359–365.
- Ruchkin, D. S., Sutton, S., & Tueting, P. (1975). Emitted and evoked P300 potentials and variation in stimulus probability. *Psychophysiology*, 12, 592–595.
- Schapkin, S. A., Falkenstein, M., Marks, A., & Griefahn, B. (2006). Executive brain functions after exposure to nocturnal traffic effects of task difficulty and sleep quality. *European Journal of Applied Physiology*, 96, 693–702.
- Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, 21, 1–54.
- Siakaluk, P. D., Buchanan, L., & Westbury, C. (2003). The effect of semantic distance in yes/no and go/no-go semantic categorization tasks. *Memory and Cognition*, 31, 100–113.
- Simpson, A., & Riggs, K. J. (2006). Conditions under which children experience inhibitory difficulty with a "button-press" go/no-go task. *Journal of Experimental Child Psychology*, 94, 18–26.
- Simson, R., Vaughan, H. G., Jr., & Ritter, W. (1977). The scalp topography of potentials in auditory and visual Go-NoGo tasks. *Electroencephalography and Clinical Neurophysiology*, 43, 864–875.
- Smith, J. L., Johnstone, S. J., & Barry, R. J. (2006). Effects of pre-stimulus processing on subsequent events in a warned Go/NoGo paradigm: Response preparation, execution and inhibition. *International Journal of Psychophysiology*, 61(2), 121–133.
- Smith, J. L., Johnstone, S. J., & Barry, R. J. (2007). Response priming in the Go/NoGo task: The N2 reflects neither inhibition nor conflict. *Clinical Neurophysiology*, 118(2), 343–355.
- Smith, J. L., Johnstone, S. J., & Barry, R. J. (2008). Movement-related potentials in the Go/NoGo task: The P3 reflects both cognitive and motor inhibition. *Clinical Neurophysiology*, 119, 704–714.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 1461–1492.
- Stadler, W., Klimesch, W., Pouthas, V., & Ragot, R. (2006). Differential effects of the stimulus sequence of CNV and P300. *Brain Research*, 1123, 157–167.
- Thorpe, S. J., & Fize, D. (1996). Speed of processing in the human visual system. *Nature*, 381, 520–522.
- VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects. *Perception*, 30, 655–668.
- van Veen, V., & Carter, S. C. (2002). The timing of action-monitoring processes in the anterior cingulate cortex. *Journal of Cognitive Neuroscience*, 14, 593–602.
- Weintraub, S. (2000). Neuropsychological assessment of mental state. In M. M. Mesulam (Ed.), *Principles of behavioral and cognitive neurology* (pp. 121–173). New York: Oxford University Press.



Contents lists available at ScienceDirect

## International Journal of Psychophysiology

journal homepage: [www.elsevier.com/locate/ijpsycho](http://www.elsevier.com/locate/ijpsycho)

## Frontal theta and alpha power and coherence changes are modulated by semantic complexity in Go/NoGo tasks

Matthew R. Brier<sup>a</sup>, Thomas C. Ferree<sup>c</sup>, Mandy J. Maguire<sup>a,b</sup>, Patricia Moore<sup>a</sup>, Jeffrey Spence<sup>e</sup>, Gail D. Tillman<sup>a</sup>, John Hart, Jr.<sup>a,d,\*</sup>, Michael A. Kraut<sup>f</sup><sup>a</sup> Berman Laboratory for Learning and Memory, Center for BrainHealth, The University of Texas at Dallas, United States<sup>b</sup> Callier Center for Communication Disorders, The University of Texas at Dallas, United States<sup>c</sup> Department of Radiology, University of Texas Southwestern Medical Center, United States<sup>d</sup> Department of Neurology, University of Texas Southwestern Medical Center, United States<sup>e</sup> Department of Internal Medicine, University of Texas Southwestern Medical Center, United States<sup>f</sup> Department of Radiology, The Johns Hopkins School of Medicine, United States

## ARTICLE INFO

## Article history:

Received 29 January 2010

Received in revised form 26 July 2010

Accepted 30 July 2010

Available online xxxx

## Keywords:

Response inhibition

EEG

Categorization

## ABSTRACT

To study the interactions between semantic processing and motor response inhibition, we recorded scalp EEG as subjects performed a series of Go/NoGo response inhibition tasks whose response criteria depended on different levels of semantic processing. Three different tasks were used. The first required the subject to make a Go/NoGo decision based on pictures of one particular car or one particular dog. The second used pictures of different types of cars and of dogs, and the final task used stimuli that ranged across multiple types of objects and animals. We found that the theta-band EEG power recorded during the NoGo response was attenuated as a function of semantic complexity while the peak latency was delayed in only the most complex category task. Further, frontal alpha-band desynchronization was strongest for the simplest task and remained close to baseline for the other tasks. Finally, there was significant theta-band coherence between the frontal pole and pre-SMA for the NoGo conditions across tasks, which was not found in the Go trials. These findings provide information about how more rostral frontal regions interact with the pre-SMA during response inhibition across different stimuli and task demands: specifically, level of processing affects latency, difficulty affects amplitude, and coherence is affected by whether the decision is Go or NoGo.

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

Specific types of cognitive functions or underlying brain states have long been associated with certain scalp-recordable EEG phenomena (Nunez, 1981). For example, EEG power changes in the theta band (4–8 Hz) have been linked to language operations (Hagoort et al., 2004), movement and working memory (Kahana et al., 2001; Raghavachari et al., 2006), threat processing (Aftanas et al., 2003), and information coding (Klimesch et al., 1996), while alpha (8–12 Hz) power has been linked with cortical idling (Pfurtscheller et al., 1996). What has not been studied as extensively is the nature of the interactions of the underlying EEG phenomena when two sets of cognitive operations that have similar spatial and spectral EEG manifestations proceed contemporaneously. Two cognitive tasks that are both associated with frontal theta and alpha power are response inhibition (Kirmizi-Aslan et al., 2006) and semantic processing (Bastiaansen et al., 2005). The similarity in the spatial

and spectral distribution of these processes is interesting given how often they must work together in day-to-day situations. Often, rapid decisions must be made based on the category to which objects belong (e.g., when driving, it is important to brake for children on the road, but it is not necessary to brake for leaves on the road) and this study seeks to address how the electrical manifestations of the neural processes that underpin rapid object identification and categorization interact with those related to rapid response inhibition.

The Go/NoGo task commonly elicits a frontal theta increase (Kirmizi-Aslan et al., 2006; Yamanaka and Yamamoto, 2010) that is accentuated in the NoGo compared to the Go condition. Go/NoGo tasks involve a subject being presented a series of stimuli, a majority of which mandate a 'Go' response (e.g., subject presses a button) and a minority that necessitate a 'NoGo' response (e.g., subject does not press a button). Visually driven Go/NoGo tasks involve visual processing, attention, and the decision to either inhibit or to execute a motor action. However, inhibition is unique to the NoGo condition and this has led to the proposition that inhibition is responsible for the theta increase. Event-related potential (ERP) studies of response inhibition have probed how task difficulty and related variables affect the transient brain responses: for instance, the amplitude of the N2

\* Corresponding author. Center for BrainHealth, 2200 W Mockingbird Lane, Dallas, TX 75235, United States.

E-mail address: [jhart@utdallas.edu](mailto:jhart@utdallas.edu) (J. Hart.).

ERP component depends on the level of perceptual similarity between the two stimuli (Nieuwenhuis et al., 2004). Previous investigations have shown that stimulus degradation affecting perceptual differentiation in Go/NoGo tasks can modulate ERP responses, with the least degraded stimuli eliciting the largest ERP response (Kok, 1986). Beyond perceptual distinctions, it was found that the level of semantic abstractness of Go and NoGo stimuli type (i.e., cars and dogs versus objects and animals) influenced the frontal N2/P3 complex (Maguire et al., 2009). Maguire et al. found that the abstractness of object identification and categorization influenced only the P3 NoGo ERP response, leading to the conclusion that the neural loci that mediate each of the major components (semantic and response inhibition) of the task were interacting in the performance of the task. It is intuitive that time–frequency analysis and ERP analysis may sometimes provide different windows into the same phenomenon (Luu et al., 2004). In our task, however, how ERP and time–frequency findings relate, and if this relationship changes with task demand, is unknown. Studying the time–frequency properties of a process that has a known influence on the P3 NoGo ERP response may clarify whether the same relationships hold between task parameters and the spectral, temporal and spatial properties of EEG.

Similar to theta power, alpha-band power changes are also often associated with inhibitory processes (for review, Klimesch et al., 2007). Specifically, decreases in alpha power from baseline are thought to represent cortical involvement (van Winsun et al., 1984) while high levels of alpha power purportedly reflect cortical idling (Pfurtscheller et al., 1996). Alpha power has been shown to increase when inhibiting a memory (Freunberger et al., 2009), performing top-down control (von Stein et al., 2000), and most relevant to this study, during the inhibition of a motor movement (Hummel et al., 2002). Those studies found increased alpha power when inhibition was required. These previous studies motivate the present study of the interactions between theta and alpha power in the process of response inhibition.

Some cognitive tasks that involve semantic processing elicit electrical responses that are spatially and spectrally similar to those detected with the Go/NoGo paradigm. Frontal theta power is known to be modulated by semantic violations in sentence processing (Hald et al., 2006) and category judgment (Brickman et al., 2005). Further, frontal theta power changes have been reported when subjects detect errors in semantic facts (Hagoort et al., 2004), suggesting that this activity is related to accessing semantic memory stores. In disease states, it has been shown that impairments in categorization abilities are correlated with lack of theta increases (Schmiedt et al., 2005). These findings suggest that, similar to response inhibition, frontal theta is involved in semantic tasks including those associated with the processes of identification and categorization tasks, though how the two may interact is unknown. Also, it is known from developmental literature that the cognitive abilities involved in identifying perceptual and conceptual categories differ (Mandler, 2000), and how those different judgments lead to decisions remains an open question.

Through this study, we sought to address several points. First, theta and alpha power have been associated with inhibition, but not as much attention has been paid to the effects of task difficulty on the magnitude of theta-band power changes. We used varying levels of semantic complexity to modulate task difficulty in the three experiments in this study. This allowed for the investigation of theta and alpha responses to Go/NoGo tasks of varying degrees of semantic difficulty on this EEG response. We hypothesized that characterization of these spectral signatures would reveal independent markers of response inhibition and level of semantic processing (object identification and categorization) as well as an interaction wherein the level of semantic processing influences the response inhibition process. Finally, we evaluated the coherence between different brain regions to probe how different stimulus characteristics and task difficulty influence the interaction between engaged brain regions. This study

affords insight into how two quite distinct factors (response inhibition and semantic processing) interact in the brain, and provides evidence for how these interact behaviorally as well as electrophysiologically.

## 2. Methods

### 2.1. Subjects

Twenty-six subjects (14 female; 12 male) participated in this study. All were college undergraduate students between the ages of 18 and 29 years (mean: 20 years). Students were given course credit for a psychology course in exchange for participation. None of the subjects reported neurological impairments. All were right handed and gave informed consent prior to participation in accordance with the Institutional Review Board of The University of Texas at Dallas. This study was conducted according to the Good Clinical Practice Guidelines, the Declaration of Helsinki, and the U.S. Code of Federal Regulations.

### 2.2. Tasks

The subjects participated in three Go/NoGo tasks with different types of stimuli and instructions. In each task, there were 160 (80%) 'Go' stimuli, for which the subject was instructed to press a button, and 40 (20%) 'NoGo' stimuli, for which the subject was instructed to withhold a response. In each of the three tasks, stimuli were presented for 300 ms followed by a fixation point (+) for 1700 ms. All of the stimuli were black line drawings fitted to a white 600×600 pixel square.

Many of the time–frequency studies that have reported associations between theta power and either inhibitory responses or erroneous responses used a simple perceptual distinction for the Go/NoGo decisions (Yamanaka and Yamamoto, 2010). Typically, this perceptual distinction is characterized by only rudimentary perceptual complexity (e.g., red versus green) and does not probe higher-level semantic processing (for review, Simmonds et al., 2007). To investigate how different semantic processing tasks interact with response inhibition, we have developed a set of three Go/NoGo tasks that vary in difficulty based on the level of semantic processing required to respond. The most basic task consists of the presentation of a picture of one specific car in the Go condition and one specific dog in the NoGo condition. This task mandates Go/NoGo decisions made on basic object level distinctions. However, with only one stimulus in each condition, the task could potentially be performed on a perceptual basis, i.e., one could plausibly perform this task simply by looking for the dog's nose, or by focusing on one location that is dark for one stimulus and light for the other. In these cases, the participants could make the Go or NoGo decision on whether that rudimentary perceptual feature appears. In the second task, the stimuli consist of many different types of dogs and cars. This task still functions at the object representation level, but the multiple stimuli preclude the task being performed at a purely perceptual level and performance of the task now requires basic object-level identification. Still, this object identification tends to rely on perceptual features; for example, all dogs in our group of stimuli have legs and tails while all the cars have wheels and doors. The third task includes stimuli that comprise multiple different types of animals (e.g., dogs, cats, snakes, etc.) and objects (e.g., hammer, houses, etc.), with subjects not pressing for animals. This discrimination is at the semantic-category level, and is much less dependent on perceptual similarities, since the visual shapes of items both within and between the main categorical distinctions (animals versus objects) were heterogeneous.

In the first task (Single-object), the Go stimulus was a line drawing of a specific car and the NoGo stimulus was a line drawing of a specific dog. In the second task (Multiple-object), the Go stimuli included 40 different car stimuli and the NoGo stimuli included 10 different dogs,

which were each shown four times. In the third task, (Semantic-category), the Go stimuli consisted of 160 objects (40 food items, 40 cars, 20 clothing items, 20 kitchen items, 20 body parts, and 20 tools) and the NoGo stimuli comprised 40 animals of varying visual typicality. The order of the tasks across subjects was counterbalanced to avoid order effects in the analysis.

### 2.3. Data acquisition and preprocessing

Continuous EEG was recorded from a 64-electrode Neuroscan Quickcap using Neuroscan SynAmps2 amplifiers and Scan 4.3.2 software sampled at 1 kHz with impedances typically below 10 k $\Omega$ . The data were processed to remove ocular and muscle artifacts. The continuous EEG data were high-pass filtered from 0.15 Hz, and then segmented offline into 2 s epochs spanning from 500 ms before to 1500 ms after the presentation of the visual stimuli.

The data were recorded with a reference electrode located near the vertex, resulting in small amplitudes over the top of the head. To eliminate this effect, the data were re-referenced to the average potential over the entire head, which approximates the voltages relative to infinity (Nunez, 1981). In order to reduce a slight bias in the electrode based average reference (Junghofer et al., 1999), spherical splines (Perrin et al., 1989) were fitted to the data and used to compute the average (Ferree, 2006). In subjects with a small number of bad electrodes, the splines were used to interpolate those electrodes, to yield a total of 62 data channels in every subject. For the coherence analysis, the surface Laplacian transformation was applied to reduce the effects of volume conduction (Babiloni et al., 2001). While the application of the surface Laplacian does not eliminate the effects of volume conduction, it is effective at reducing the impact of such effects on coherence analysis (Srinivasan et al., 1998). The use of this transform has been investigated and theoretically and empirically validated (Nunez et al., 1997). Further, the results presented here were compared between conditions, which controls for the constant effects of volume conduction.

### 2.4. Event-related power analysis

Fourier power spectra were computed using a slight modification of the *pwelch* function implemented in Matlab (Mathworks, Inc.), applied to 0.25-s windows. Fourier estimates of the spectra were used due to its high frequency resolution and, at low frequencies, its comparable temporal resolution to wavelets and other methods. In each epoch and time window, the time series were linearly detrended and mean subtracted to reduce spectral leakage from the zero-frequency bin, cosine tapered to reduce spectral leakage generally, and zero-padded to 1-s duration to achieve 1-Hz frequency resolution. It should be noted, however, that the rapidity of the task necessitates using windows of such short duration, which obscures detection and characterization of oscillations of lower than 4 Hz. Each window was then Fourier transformed, magnitude squared, and suitably normalized to obtain the power spectral density (PSD) in units  $\mu V^2/Hz$ . The results were averaged across trials to obtain the best statistical estimate of the PSD. Only trials to which the subject responded correctly, and those without artifacts, were included.

The time-dependent PSD was estimated in 0.25-s wide windows, moving in 0.025-s steps. The time of each window was defined as the center of the nonzero data in that window. The earliest time window was  $-0.75$  s and the latest time was  $2.75$  s, because the centers of 0.25-s windows cannot reach the ends of the epoch. For each condition, the power spectrum in each moving window was averaged across all trials, to obtain the best statistical estimate of the PSD. Finally, the mean baseline power at each electrode and frequency was subtracted to highlight power differences relative to baseline (Delorme and Makeig, 2004).

In order to measure the relative phase locking of observed oscillations, the phase locking value (PLV) was calculated. The method by which this value can be calculated from Fourier spectra is detailed elsewhere (Delorme and Makeig, 2004). Briefly, PLV is a measure of the consistency of the phase of an oscillation across trials, relative to some event. If an oscillation has constant phase across trials (i.e., all the peaks and troughs align in time) the value is 1, and if the phase is random across trials the value is 0. Thus a PLV close to 1 indicates that time-domain averaging of an oscillatory signal is likely to result in those oscillations being reflected in the ERP.

### 2.5. Peak analysis

For the hypothesis-driven analysis of frontal theta power, we performed a peak picking analysis on the midline frontal electrodes (Fz, F1 and F2) in the theta band. Within a subject and condition, PSD is a three dimensional matrix spanning space, frequency, and time. First, the time series for the three electrodes of interest and the five frequencies of interest (4–8 Hz) were averaged. Then, the maximal value and latency of that value in the time period ranging from 0 s to 750 ms was computed. At the single subject level, these time courses had only a single maximum, and thus provided a robust measure of the theta response.

### 2.6. Data reduction

We applied a previous developed technique, called STAT-PCA, for reducing the dimensionality of space–time–frequency (STF) data collected in cognitive tasks (Ferree et al., 2009). More specifically, it uses an extension of that original technique that accounts for both within-subject and across-subject variance using block-design ANOVAs with subjects as the blocking factor. This technique has previously been used to identify time–frequency changes associated with known cognitive processes (Maguire et al., 2010). We extend this approach further, to include more than two task conditions. This extension is straightforward because the statistical inference relies upon ANOVA, which can have an arbitrary number of independent variables.

The present experiment has two independent variables: task and condition. The task variable has three levels (Single-object, Multiple-object, and Semantic-category), and the condition variable has two levels (Go and NoGo). The statistical model can be written  $P_{ctsk} = \mu + C_c + T_t + (CT)_{ct} + B_s + \epsilon_{ctsk}$ , where  $P$  is the observed power at a given space–time–frequency point,  $\mu$  is the general mean,  $C$  is the mean contributed by condition  $c$ ,  $T$  is the mean contributed by task  $t$ ,  $k$  is the trial number, and  $B$  is the blocking factor accounting for the mean contributed by subject  $s$ . It should be noted that the task and condition are fixed effects, while the subjects and trials are random effects. The ANOVA returns (for each space–time–frequency point) three  $F$  statistics: main effect of task, main effect of condition, and interaction of task and condition. These values were input to sequential PCA. Thus the statistical step of STAT-PCA, accomplished by ANOVA, reduces the dimensionality as follows:

$$P(r, f, \tau, t, c, s, k) \xrightarrow{\text{ANOVA}} \begin{cases} F_{\text{condition}}(r, f, \tau) \\ F_{\text{task}}(r, f, \tau) \\ F_{\text{both}}(r, f, \tau) \end{cases}$$

All  $F$  values that resulted in  $p$  values above the threshold ( $p > 0.05$ ) were set to 0.

The resulting  $F$  statistics have dimensions of space (electrodes) by frequency by time matrix, and were reduced with sequential PCA. In keeping with previous work (Ferree et al., 2009), the frequency dimension was analyzed first. The matrix was reshaped such that the columns indexed frequency and the rows indexed time and space. PCA returned a set of eigenvalues and eigenvectors (factor loadings). In order to determine the number of factors to retain, Parallel Analysis



**Table 1**

The mean accuracy and reaction time are shown; standard deviations are in parentheses.

Task	Go correct	NoGo correct	Go RT (ms)
Single-object	93.9%(17.6)	82.7%(18.9)	293.7(5.9)
Multiple-object	93.1%(17.9)	83.3%(18.6)	307.7(5.8)
Semantic-category	90.8%(18.2)	82.5%(18.5)	361.1(10.7)

was employed (Horn, 1965; Ferree et al., 2009). This estimates the eigenvalues of a null matrix, comprised of the same elements in random positions, and only eigenvalues above the 95% confidence intervals of the null distribution are retained. The retained eigenvectors were rotated to meet the Varimax criterion. In order to compute the corresponding factor scores, the original dataset was projected onto these eigenvectors. Each factor score was reshaped, such that the columns indexed electrodes and the rows indexed time points. A second PCA was performed on this matrix, including similar steps for factor retention and rotation. The resulting eigenvectors represented topographies, and their corresponding factor scores represented their time courses. The end result is a set of factor “triplets”, comprised of spectral, spatial, and temporal factors derived through PCA.

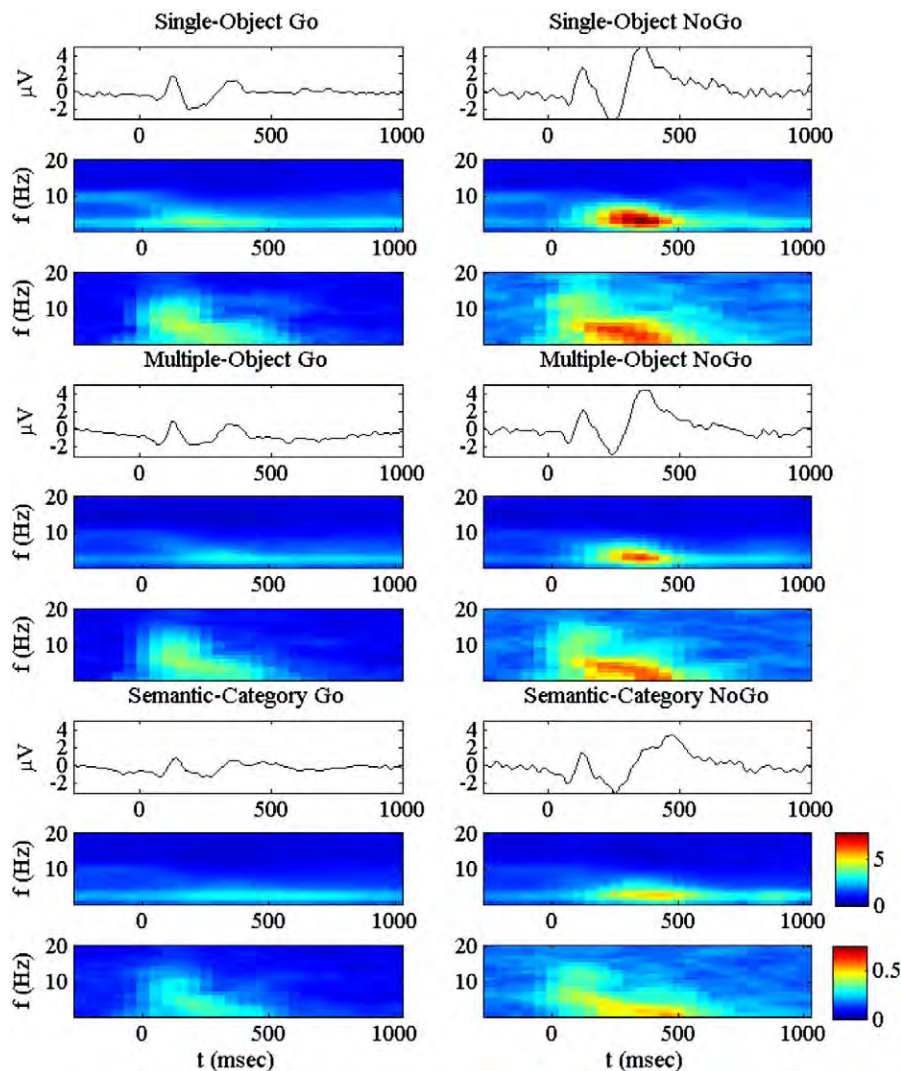
## 2.7. Event-related coherence analysis

Coherence was computed using the short-window Fourier transform, applied to 0.25-s windows. In each window, the time series was linearly detrended to reduce spectral leakage from the zero-frequency bin, cosine tapered to reduce spectral leakage generally, zero-padded to 1-s duration, and Fourier transformed to achieve 1-Hz frequency resolution. To form the numerator of coherence, the cross-spectrum of two channels was computed by multiplying their Fourier transforms, averaging across trials, and computing the magnitude squared. To form the denominator (i.e., to normalize for power differences), the power of each channel was computed by averaging across trials.

## 3. Results

### 3.1. Behavioral data

The behavioral data for this task have been published previously (Maguire et al., 2009) and are summarized here (see Table 1). Analysis of the response time (RT) showed that the Semantic-category task had a significantly longer RT than the other two tasks. Accuracy did



**Fig. 1.** Time–frequency and PLV spectrograms derived from the frontal electrodes (Fz, F1, and F2) as well as averaged ERPs at the same electrodes. Go (left column) and NoGo (right column) responses are shown for each task (rows). The first, fourth, and seventh rows show the ERP, the second, fifth, and eighth rows show the power (PSD), and the third, sixth and ninth rows show the PLV.

not differ significantly across tasks. This suggests that the Semantic-category task is inherently more difficult.

### 3.2. Event-related power: hypothesis-driven

Previous research has shown that medial frontal regions are involved in response inhibition (Yamanaka and Yamamoto, 2010) and that these areas can respond differentially based on the difficulty of the task (Maguire et al., 2009). In order to test the hypothesis that semantic complexity would influence these frontal regions in our response inhibition task we performed a task by condition analysis on the time–frequency data from this region. We thus evaluated midline and immediately para-midline frontal electrodes (Fz, F1, F2) and obtained the time–frequency spectrograms (Fig. 1). For each panel in Fig. 1, the data presented represent the average across all subjects and across the three electrodes of interest (Fz, F1, and F2). These electrodes were chosen based on previous research in this task (Maguire et al., 2009) as well as response inhibition in general (Yamanaka and Yamamoto, 2010). For comparison, the ERP time courses are also shown, but for a statistical treatment of those data, see Maguire et al. (2009). The Go conditions for all three tasks showed a small increase in the theta (4–8 Hz) band while the NoGo conditions showed a larger response (There is also activity below 4 Hz, but that activity does not differ across conditions,  $p > 0.25$ ). In addition, the NoGo increase in theta power was progressively attenuated as the semantic complexity of the stimuli/task increased (i.e., amplitude during the Single-object NoGo epochs > Multiple-object Identification > Semantic-category). Each single subject's peak theta power and latency were separately submitted to task by condition repeated measures ANOVA. The amplitude showed a main effect of task (Single-object, Multiple-object Identification, Semantic-category)  $F(2,25) = 13.19$ ,  $p = 0.00002$ , as well as a main effect of condition (Go versus NoGo)  $F(1,25) = 16.50$ ,  $p = 0.0004$ . However, there was also a condition by task interaction,  $F(2,25) = 8.9441$ ,  $p = 0.0005$ .

Fig. 2 shows the interaction in theta power between task and condition. To investigate this interaction, we performed one-way repeated measures ANOVA by splitting the previous two-way repeated measures ANOVA according to condition. The Go condition showed no significant task effect ( $F(2,25) = 0.954$ ,  $p = 0.39$ ), but the NoGo ANOVA did ( $F(2,25) = 10.551$ ,  $p = 0.0002$ ). Post-hoc  $t$ -tests on the NoGo condition theta power revealed that Single-object responses were significantly different from (larger than) both Multiple-object ( $t = 2.53$ ,  $p = 0.018$ ) and Semantic-category ( $t = 3.80$ ,  $p = 0.0008$ ), and that theta power measured in the Multiple-object task was significantly different from (larger than) that in the Semantic-category task ( $t = 2.7991$ ,  $p = 0.0097$ ). Thus, the observed differences

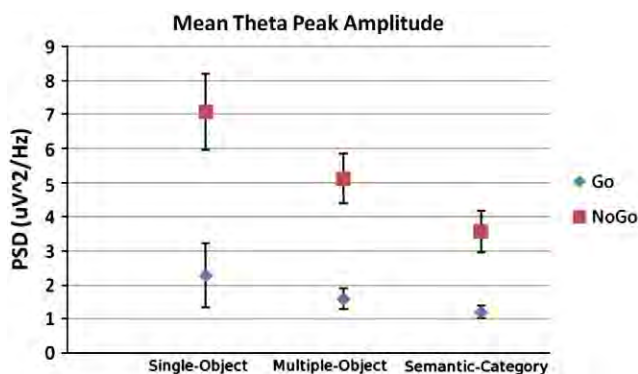


Fig. 2. Summary of theta PSD in each task and condition. The first column depicts the theta amplitudes in the Go and in the NoGo conditions for the Single-object task, second column shows the amplitudes from the Multiple-object task, and the third column from Semantic-category task. The figure shows no significant differences in the Go condition across tasks and a systematic attenuation of the theta power increase across tasks in the NoGo condition. Error bars represent one Standard Error.

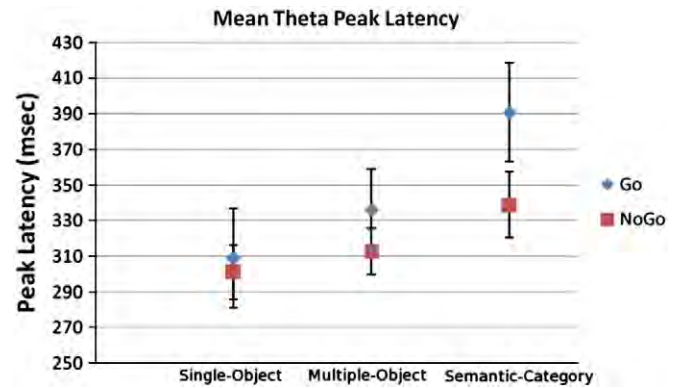


Fig. 3. The peak theta latency depicted using the same format as Fig. 2. The results show similar latencies for the Single-object and Multiple-object tasks and a significant increase in latency for the Semantic-category task.

in theta-band power increases across tasks occurred primarily in the NoGo condition.

The latency of the peak theta activity was also submitted to a task by condition repeated measures ANOVA. There was a significant effect of task,  $F(2,25) = 4.88$ ,  $p = 0.0116$  (see Fig. 3). Post-hoc  $t$ -tests revealed that the peak latency of the power increase during performance of the Semantic-category task was significantly later than it was for the both Single-object ( $t = 2.79$ ,  $p = 0.009$ ) and Multiple-object tasks ( $t = 2.78$ ,  $p = 0.01$ ), but that the peak latencies for the Single-object and Multiple-object tasks were not significantly different from each other ( $t = 0.87$ ,  $p = 0.39$ ).

### 3.3. Phase locking value (PLV) and relationship to ERP

The PLV is shown in Fig. 1 for each task and condition. Roughly speaking, the “hot” spots in power are also the “hot” spots in the PLV. Closer inspection reveals interesting details; however, the following discussion is limited to the red and yellow areas of the graphs. In the theta band, the changes in power and PLV share roughly the same time interval, for all tasks and conditions. This predicts a tight link between theta power and the corresponding ERP. In the delta band, particularly in the NoGo condition, the PLV changes appeared to start and end slightly later than the power changes. This too is reflected in the ERP, as a slowing of the oscillation followed by a more gradual decline toward baseline.

In order to investigate further the connection between the theta power changes and the ERP, we performed a correlation analysis between the peak theta power changes and the peak ERP amplitudes, for each task and condition, with subjects as samples. Fig. 4 shows a

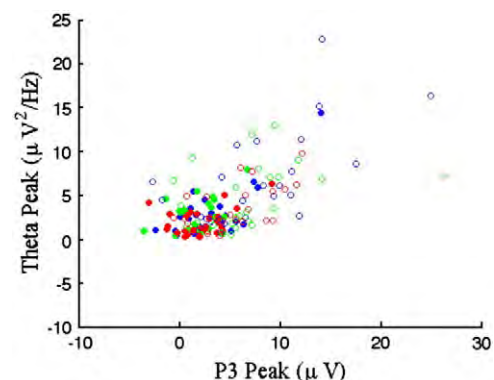
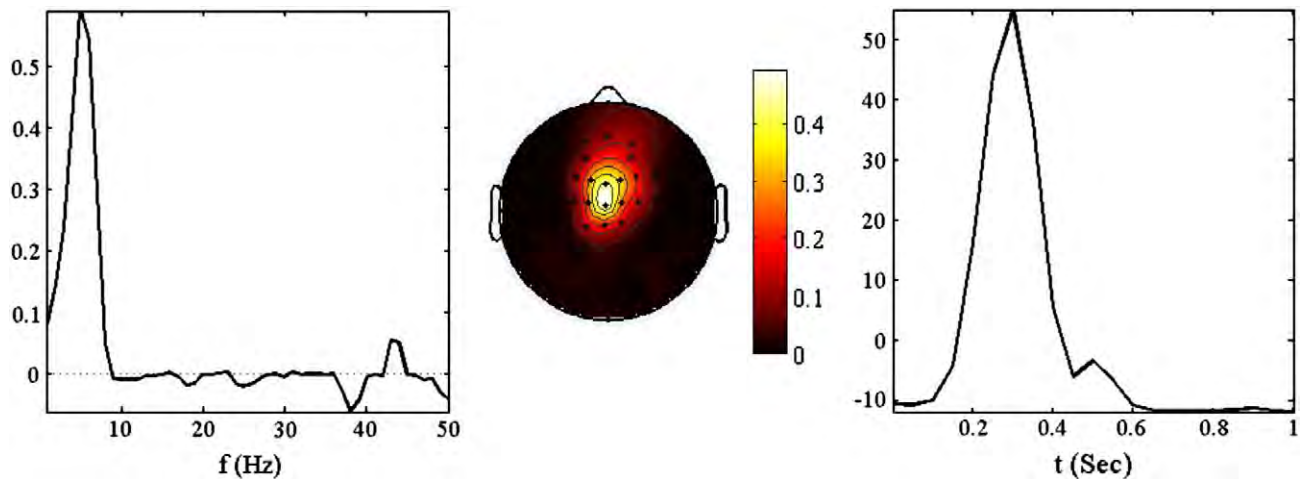


Fig. 4. The correlation between P3 amplitude and theta power amplitude is shown. The Single-object task is shown in blue, the Multiple-object task is shown in green, and the Semantic-category task is shown in red. Solid marks indicate the Go condition while hollow marks indicate the NoGo condition. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 5.** STAT-PCA result for the interaction of task and condition. The spectrum (left panel) is loaded in the theta band, the peak (center panel) is topographically centered to the midline frontal regions, and the time course (right panel) is maximal around 0.3 s. These distributions in frequency, space, and time are identical to what was found in the hypothesis-driven analysis.

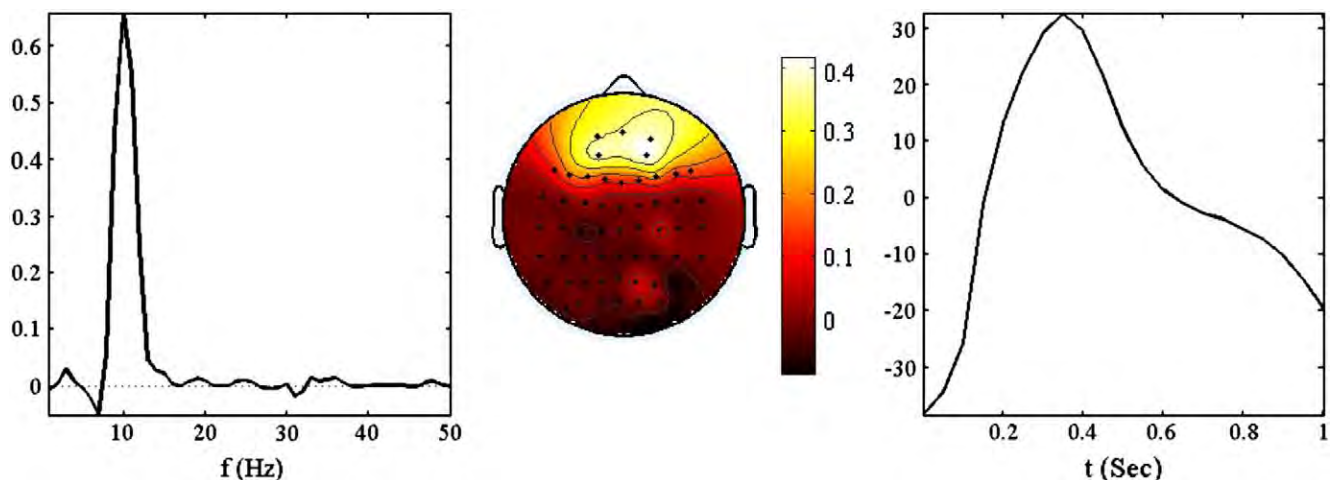
scatter plot of these quantities. Correlation analysis showed that these values were highly correlated in all tasks and conditions ( $r > 0.5$ ,  $p < 0.005$ ; Fig. 4). In summary, these results show that the ERP in this task is closely associated with increases in theta power. In discussing this relationship, it is common to attempt to distinguish between evoked responses, in which the ERP is associated with increased power that adds to the background EEG, and induced responses, in which the ERP is due to phase resetting of pre-existing power (Sauseng et al., 2007). In the results presented here, the ERP differences are explained to a great degree by evoked activity, although the present analysis cannot rule out the possibility of some phase resetting of ongoing activity as well (Kalcher and Pfurtscheller, 1995). In contrast, to the extent that delta-band changes were visible in the PLV, but not visible in the power, these delta-band changes would be characterized best as induced responses due entirely to phase resetting.

### 3.4. Event-related power: data driven

The first PCA decomposition was performed on the  $F$  matrix representing the task by condition interaction. This yielded one significant factor (Fig. 5). This factor was loaded in the theta band,

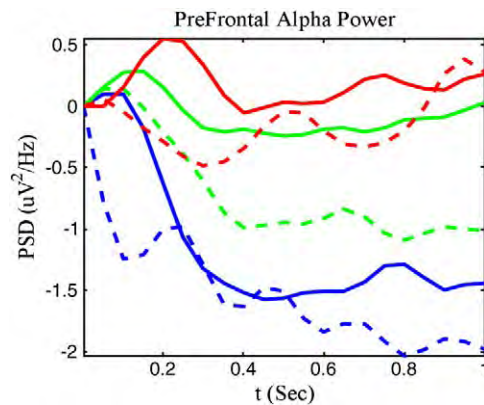
over midline frontal regions, at 0.3 s. This was the same spectrum, spatial distribution, and time course found in the hypothesis-driven analysis so no further investigation was performed on this factor. This analysis, while revealing no new interactions of task and condition, serves to validate the *a priori* hypothesis-driven approach.

The second PCA analysis was performed on the  $F$  matrix representing main effects of task. This yielded one component. The component (Fig. 6) was loaded in the alpha band ( $\sim 10$  Hz) over midline prefrontal regions. The time course increased from baseline at task onset and began to decline around 600 ms. The individual time series are shown in Fig. 7, but as this is a main effect of task there is no significant difference between the conditions. A summary figure is shown (Fig. 8). The Single-object (blue) task showed a notable decline from baseline, the Multiple-object (green) task showed a smaller decline, and the Semantic-category (red) task showed no change from baseline. In summary, the alpha power shows a decrease from baseline in all tasks, but the Single-object task shows the greatest decrease and the Semantic-category task shows the smallest decrease, with the Multiple-object task exhibiting an intermediate degree of reduction in power. Post-hoc tests reveal that the Single-object task was significantly different than the other two tasks ( $t = 1.95$ ,  $p = 0.008$ ), but Multiple-object and Semantic-category tasks were



**Fig. 6.** STAT-PCA result for the main effect of task. The spectrum (left panel) is loaded in the alpha band, the topography of the peak (center panel) is over the midline prefrontal regions, and the time course (right panel) is maximal around 0.4 s but remains elevated throughout the task.





**Fig. 7.** Time courses of the prefrontal alpha response for the Go (solid) and NoGo (dashed) conditions in the Single-object (blue), Multiple-object (green), and Semantic-category (red) task. There was no significant difference between Go and NoGo conditions, but there was a significant effect of task. The time courses show a drop in alpha power for the Single-object task but a smaller drop in power for the Multiple-object and Semantic-category task. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

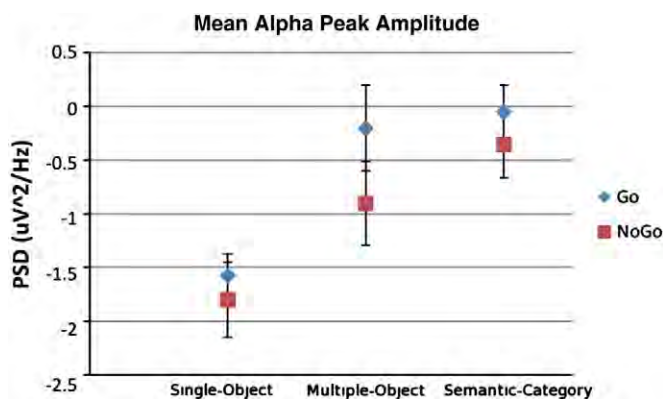
not significantly different ( $p > 0.1$ ) from one another. The main effect of condition yielded no components that were not accounted for in the interaction.

### 3.5. Event-related coherence

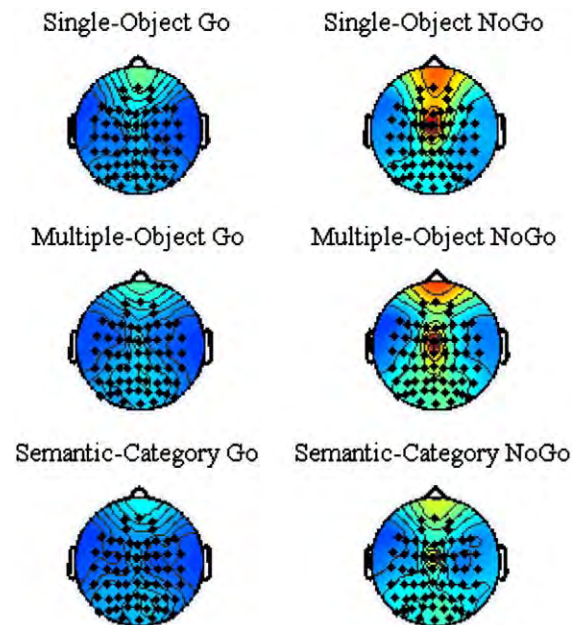
To investigate the topography of the theta activation, the theta-band EEG power at the time of the peak amplitude was plotted topographically (Fig. 9). This demonstrated two distinct local maxima for the theta power increase, one with topography over frontal pole and the other over approximate midline pre-SMA areas. Given these two circumscribed regions, we conducted a coherence analysis between these two electrodes, the results of which are shown in Fig. 10. The single subjects' peak theta coherence amplitudes and latencies were submitted to task by condition repeated measures ANOVAs separately. The magnitude of the coherence was found to be significantly higher in the NoGo versus Go condition,  $F(1,25) = 132.7946$ ,  $p < 0.0001$ . There was no effect of task or an interaction for the magnitude of the coherence. The latency of the peak coherence showed no significant effects,  $F < 1$ .

## 4. Discussion

This study demonstrates significant differences in the amplitude of the theta power associated with the NoGo response across tasks, with



**Fig. 8.** The mean alpha amplitude depicted using the same format as Fig. 2. The results show a large decrease in power following stimulus presentation for the Single-object task, and intermediate drop in power for the Multiple-object task, and a negligible drop in power for the Semantic-category task.



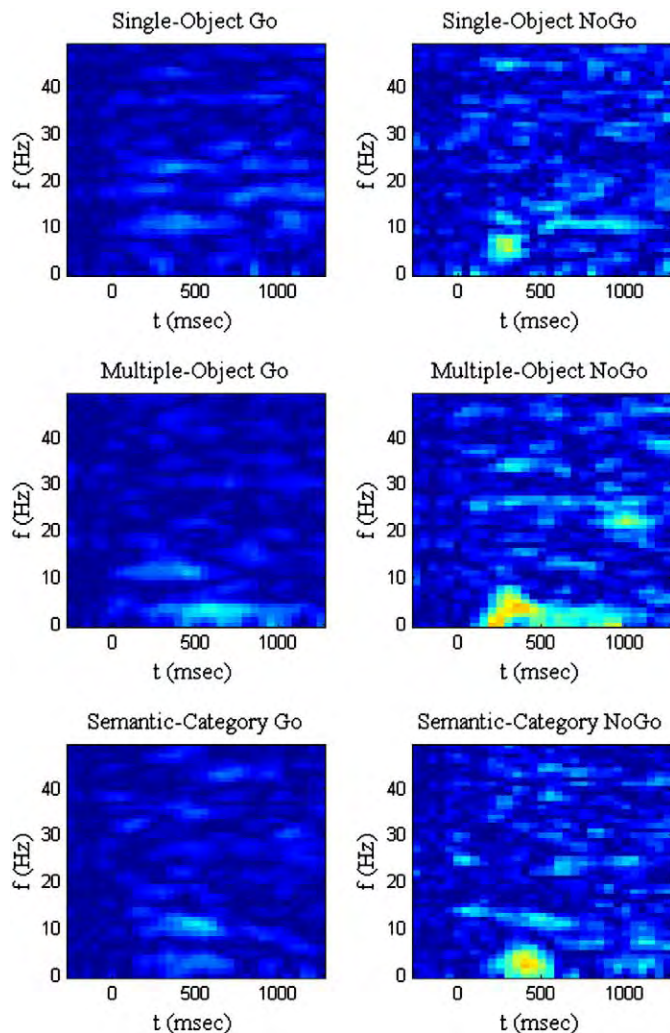
**Fig. 9.** The topography of theta power at 0.3 s in all tasks and conditions. Note that there are two local maxima, one over FCz and one over FPz, and the attenuation of theta power across tasks.

attenuated increases in theta power associated with increases in semantic task/stimuli complexity. The presence of theta power increases with NoGo responses is consistent with previous reports that theta power changes are generally associated with inhibitory cognitive processes (Yamanaka and Yamamoto, 2010). The localization of these theta changes to the midline frontal regions is also consistent with previous studies suggesting that successful NoGo trials engage a frontal cortical-subcortical circuit that mediates effective inhibition of the motor response (Aron and Poldrack, 2006; Mostofsky et al., 2003). Second, the peak latency of the theta power changes differs between tasks, with the Semantic-category task-related theta power peaking at a significantly later time than the other two tasks. Third, there were two distinct loci of theta activation found in midline frontal regions that showed increased coherence in the NoGo condition compared to the Go condition. Finally, the exploratory STAT-PCA confirmed the theta-band effect found in the *a priori* investigation using data-driven analysis, and also revealed an alpha power effect in frontal regions that was not considered in the *a priori* investigations.

### 4.1. Theta power amplitude

There was an increase in theta-band power, especially during the NoGo trials in each of our experiments, all incorporating varying degrees of both semantic processing and motor response inhibition components. The attenuation of this power increases with increasing semantic processing complexity, however, is somewhat at odds with other studies. For example, theta power has been found to increase in tasks requiring increased attention, task switching, executive functioning, and/or working memory (for review, Kahana et al., 2001; also, see Gevins et al., 1997; Burgess and Gruzeliier, 2000), and in semantic processing (Klimesch et al., 1994). In the context of our tasks, increased task difficulty and semantic complexity are intertwined, as indexed by the degree of diversity of the stimuli in the three tasks, and evidenced by the increase in reaction times proceeding from the Single-object through to the Semantic-category tasks. In addition to attenuated increases in theta power, the latency of the peak of the theta power increase grew in the NoGo trials when comparing the Semantic-category task with the two other tasks. Further, visual





**Fig. 10.** Coherence spectrograms between FCz and FPz. There is no increase in theta coherence in the Go condition but there is a significant increase in the NoGo condition around 0.3 s. There is no significant difference across tasks.

inspection of the temporal profiles of the power curves for each of the tasks revealed that, with the NoGo trials during the Semantic-category task, the curve was broader than during the other two tasks. The fact that this progression of changes is only evident during the NoGo trials indicates that the phenomenon we are observing relates predominantly to response inhibition, as mediated by semantic complexity and level of processing, *i.e.*, not semantic factors. We suggest that, if there were a prominent or isolable role of semantic complexity/level of processing *per se*, there would likely also be changes during the Go trials.

This peak widening and height reduction may reflect temporal dispersion of semantic-related inputs to the dorsomedial frontal lobes. Such dispersion could be due to inputs from the greater number of, and disparate distances to, locations in the brain required to execute the Semantic-category task. More specifically, in the most complex task, the dorsomedial frontal lobes might have to receive, process and synthesize inputs from sites mediating access to semantic memory information from multiple categories of items, while in the first two (Single-object and Multiple-object) tasks, the inputs might be coming from the same two sites or groups of sites. The smaller but still recordable dorsomedial and more rostral frontal responses with the Go trials in each task presumably reflect a combination of the object synthesis/identification role that has been attributed to the dorsomedial frontal lobes (Crosson et al., 1999; Kraut et al., 2002) and

the pre-movement activity that normally occurs in supplementary and other non-primary motor regions. It should be noted that this pattern of latency differences we detected in time-varying theta power replicates previous findings on the P3 ERP component in this task (Maguire et al., 2009) suggesting that, at least in the context of this task, the EEG and ERP measures index the same or temporally closely linked cortical processes.

In these Go/NoGo tasks, the NoGo stimuli were rare (20% of trials) compared to the Go stimuli (80% of trials), creating a potential confound between the NoGo response and an oddball response (Nieuwenhuis et al., 2003). Previous studies of oddball responses indicate that the oddball stimulus can evoke delta and theta power increases in post-stimulus time epochs similar to those reported here (ar-Eroglu et al., 1992). The differences in theta power across tasks reported here, however, were not likely due to the oddball effect, because we found reduced theta power with increasing task complexity (*i.e.*, tasks with fewer repeated stimuli), while previous studies found a larger response for more rare stimuli in the oddball task (Gruber and Muller, 2005; Courchesne, 1978).

Differences in task difficulty also modulate the oddball response. Other authors have reported that increasing task difficulty increases the frontal P3 amplitude (Hagen et al., 2006; Sawaki and Katayama, 2009), which is similar to the presentation frequency effect. This is inconsistent with our findings, where increased task difficulty reduced frontal theta power, and tracked roughly P3 amplitude. Thus, although similar power frequencies are reported in both tasks, the opposite direction of the effect suggests that our NoGo effects are not simply the oddball effect.

#### 4.2. Theta power topography

Our data exhibit two distinct theta power maxima in the midline frontal region, one close to the frontal pole, and the other in the pre-SMA area. The locations of these signal peaks did not vary across our tasks, although the neural activity in both the SMA and the polar regions is modulated by semantic difficulty. fMRI studies of response inhibition identify many different loci of activation (Rubia et al., 2001) that vary with the types of stimuli used and the semantic judgments upon which the response inhibition is contingent; however, lesion studies have provided insight into which areas are essential for performance of inhibition tasks. Lesions of the frontal pole have been shown to increase reaction times substantially in Go/NoGo tasks (Picton et al., 2007), and lesions of pre-SMA areas have been shown to increase the number of false-positives in the NoGo condition (Drewe, 1975; Picton et al., 2007). These lesion-related findings support the concept that the midline frontal polar regions facilitate the task but are not essential to its performance, whereas the midline pre-SMA area is more immediately responsible for inhibiting the motor action. While lesion studies do not directly specify the cognitive operations underlying the neural activity in the frontal polar region, functional imaging studies have implicated this region with action selection, response inhibition, and performance monitoring (Ridderinkhof et al., 2004), all of which are plausible in the setting of the three tasks reported here. Our findings support the hypothesis that in the performance of motor inhibition tasks in general, the frontal polar regions monitor and select the action of inhibition, while pre-SMA is more directly engaged in inhibiting the motor response, with both the difficulty and level of semantic processing modulating the neural activity in these regions.

#### 4.3. Theta coherence

EEG coherence is a measure of dynamic interaction between neural structures and has often been used in the study of cognition (Sauseng et al., 2005; Rappelsberger and Petsche, 1988; Weiss and Mueller, 2003). Coherence analysis indicated increased theta

interactions between the frontal polar and the pre-SMA regions in the NoGo versus Go conditions, with no significant differences across the tasks. This supports the notion that the frontal pole interacts with and modulates the pre-SMA, which has been implicated in both motor inhibition (Rubia et al., 2001) and semantic processing (Kraut et al., 2002; Crosson et al., 1999). A finding of increased synchrony between these brain regions in the NoGo condition is entirely consistent with the hypothesis of greater motor control required to inhibit a response.

#### 4.4. Relationship between theta and ERP measures

The results presented here sought to relate both ERP and time-frequency measures. One measure that facilitates this is the PLV. The theta power increases in power were concurrent with increases in PLV, which predicts that theta-band oscillations should be manifest in the ERP. This prediction was confirmed by the high correlation between the ERP and power measures, and suggests that the theta power increases reported here are strongly linked to the stimulus and highly consistent across trials.

#### 4.5. Alpha power

Beyond its general interpretation as an idling rhythm (Pfurtscheller et al., 1996), alpha power has been associated specifically with semantic processing (Klimesch et al., 1997) as well as inhibition (Klimesch et al., 2007) though controversy remains (Cooper et al., 2003). Alpha power in midline prefrontal electrodes showed attenuation in all three tasks, with the greatest attenuation in the Single-object task, and no effect of condition. This finding differs from the finding for theta power in three ways. First, alpha power was attenuated, while theta power was enhanced in the NoGo condition generally. Second, the alpha effect occurred in both conditions equally, while the theta effect was greater in the NoGo condition, further suggesting theta power association with the inhibitory component in these tasks. Third, the alpha effect was similar in the two more difficult tasks, both of which require some type of semantic classification. The Single-object task is cognitively simpler, and the relative lack of alpha suppression in the more demanding Multiple-object and Semantic-category tasks is counter to interpretation of alpha power changes in these tasks as an index of cortical idling (Pfurtscheller et al., 1996). This lack of alpha suppression may reflect an as-of-yet unclear interaction between aspects of the semantically more complex tasks and response inhibition. Regardless of the direction of this effect, there is no differentiation between the Go and NoGo responses, suggesting that alpha power still indexes overall task difficulty in some regard. However, there is evidence that the inability to suppress alpha power (following traumatic brain injury) results in poor performance in a response inhibition task (Roche et al., 2004), which challenges the notion that alpha power is only involved in the semantic portion of this task.

#### 4.6. Summary

This study demonstrates that in inhibition tasks, semantic complexity attenuates the amplitude of theta power, while the engagement of semantic categorization versus object identification affects the latency of the theta response. These theta power changes are noted in two regions for all of the tasks – frontal pole and midline pre-SMA. Activity in the pre-SMA and frontal poles is closely related to task difficulty (measured by RTs) and may be directly proportional to it, in that there is increased theta power in the tasks requiring lower levels of semantic processing. This study provides some information of how semantic complexity modulates response inhibition, at least at a scalp-observable electrophysiological level, which supplements previous work on the perceptual modulation of response inhibition (Nieuwenhuis et al., 2004). The task invoking a semantic categoriza-

tion judgment had a delayed peak latency of theta power changes compared to the two object identification level tasks, demonstrating that level of semantic processing modulates the time course of theta power. The frontal polar and pre-SMA areas showed increased theta synchronization for the NoGo versus Go condition, which was of about the same magnitude across all tasks and theta amplitude. Frontal lobe alpha power showed no distinction between Go and NoGo components of any task, although there was less suppression for the two more difficult tasks. Compared to previous research focusing on response inhibition of perceptual stimulus characteristics (e.g., red-green, triangle-square, direction of arrow, etc.), we systematically varied the stimulus characteristics and task by levels of semantic conceptual processing and detected multiple modulations of the response inhibition circuit.

#### References

- Aftanas, L.I., Pavlov, S.V., Reva, N.V., Varalimov, A.A., 2003. Trait anxiety impact on EEG theta band power changes during appraisal of threatening and pleasant visual stimuli. *International Journal of Psychophysiology* 50, 205–212.
- Ar-Eroglu, C., Basar, E., Demiralp, T., Schürmann, M., 1992. P300-response: possible psychophysiological correlates in delta and theta frequency channels. A review. *International Journal of Psychophysiology* 13, 161–179.
- Aron, A.R., Poldrack, R.A., 2006. Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus. *The Journal of Neuroscience* 26, 2424–2433.
- Babiloni, F., Cincotti, F., Carducci, F., Rossini, P.M., Babiloni, C., 2001. Spatial enhancement of EEG data by surface Laplacian estimation: the use of magnetic resonance imaging based head models. *Clinical Neurophysiology* 112, 724–727.
- Bastiaansen, M.C.M., Linden, M., Keurs, M., Dijkstra, T., Hagoort, P., 2005. Theta responses are involved in lexical-semantic retrieval during language processing. *Journal of Cognitive Neuroscience* 17, 530–541.
- Brickman, A.M., Paul, R.H., Cohen, R.A., Williams, L.M., MacGregor, K.L., Jefferson, A.L., Tate, D.F., Gunstad, J., Gordon, E., 2005. Category and letter verbal fluency across the adult lifespan: relationship to EEG theta power. *Archives of Clinical Neuropsychology* 20, 561–573.
- Burgess, A.P., Gruzelić, J.H., 2000. Short duration power changes in the EEG during recognition memory for words and faces. *Psychophysiology* 37, 596–606.
- Cooper, N.R., Croft, R.J., Dominey, S.J., Burgess, A.P., Gruzelić, J.H., 2003. Paradox lost? Exploring the role of alpha oscillations during externally vs internally directed attention and the implications for idling and inhibition hypotheses. *International Journal of Psychophysiology* 47, 65–74.
- Courchesne, E., 1978. Changes in P3 waves with event repetition: long-term effects on scalp distribution and amplitude. *Electroencephalography and Clinical Neurophysiology* 45, 754–766.
- Crosson, B., Sadek, J.R., Bobholz, J.A., Gökçay, D., Mohr, C.M., Leonard, C.M., Maron, L., Auerbach, E.J., Browd, S.R., Freeman, A.J., Briggs, R.W., 1999. Activity in the paraculate and cingulate sulci during word generation: an fMRI study of functional anatomy. *Cerebral Cortex* 9, 307–316.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods* 134, 9–21.
- Drewe, E.A., 1975. Go-no go learning after frontal lobe lesions in humans. *Cortex* 11, 8–16.
- Ferree, T.C., 2006. Spherical splines and average referencing in scalp electroencephalography. *Brain Topography* 19, 43–52.
- Ferree, T.C., Brier, M.R., Hart, J., Kraut, M.A., 2009. Space-time-frequency analysis of EEG data using within subject statistical tests followed by sequential PCA. *Neuroimage* 45, 109–121.
- Freunberger, R., Fellinger, R., Sauseng, P., Gruber, W., Klimesch, W., 2009. Dissociation between phase-locked and nonphase-locked alpha oscillations in a working memory task. *Human Brain Mapping* 30, 3417–3425.
- Gevens, A., Smith, M.E., McEvoy, L., Yu, D., 1997. High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cerebral Cortex* 7, 374–385.
- Gruber, T., Müller, M.M., 2005. Oscillatory brain activation dissociates between associative stimulus content in a repetition priming task in the human EEG. *Cerebral Cortex* 15, 109–116.
- Hagen, G.F., Gatherwright, J.R., Lopez, B.A., Polich, J., 2006. P3a from visual stimuli: task difficulty effects. *International Journal of Psychophysiology* 59, 8–14.
- Hagoort, P., Hald, Bastiaansen, M., Petersson, K.M., 2004. Integration of word meaning and world knowledge in language comprehension. *Science* 16, 438–441.
- Hald, L.A., Bastiaansen, M.C.M., Hagoort, P., 2006. EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language* 96, 90–105.
- Horn, J.L., 1965. A rationale and test for the number of factors in factor analysis. *Psychometrika* 30, 179–185.
- Hummel, F., Andres, F., Altenmüller, E., Dichgans, J., Gerloff, C., 2002. Inhibitory control of acquired motor programmes in the human brain. *Brain* 125, 404–420.
- Junghofer, M., Elbert, T., Tucker, D.M., Braun, C., 1999. The polar average reference effect: a bias in estimating the head surface integral in EEG recording. *Clinical Neurophysiology* 110, 1149–1155.

- Kahana, M.J., Seelig, D., Madsen, J.R., 2001. Theta returns. *Current Opinions in Neurobiology* 11, 739–744.
- Kalcher, J., Pfurtscheller, G., 1995. Discrimination between phase-locked and non-phase-locked event-related EEG activity. *Electroencephalography and Clinical Neurophysiology* 94, 381–384.
- Kirmizi-Aslan, E., Bayraktaroglu, Z., Gurvit, H., Keskin, Y.H., Emre, M., Demiralp, T., 2006. Comparative analysis of event-related potentials during Go/NoGo and CPT: decomposition of electrophysiological markers of response inhibition and sustained attention. *Brain Research* 1104, 114–128.
- Klimesch, W., Schimke, H., Schwaiger, J., 1994. Episodic and semantic memory: an analysis in the EEG theta and alpha band. *Electroencephalography and Clinical Neurophysiology* 91, 428–441.
- Klimesch, W., Doppelmayr, M., Russegger, H., Pachinger, T., 1996. Theta band power in human scalp EEG and the encoding of new information. *Neuroreport* 17, 1235–1240.
- Klimesch, W., Doppelmayr, M., Pachinger, T., Russegger, H., 1997. Event-related desynchronization in the alpha band and the processing of semantic information. *Cognitive Brain Research* 6, 83–94.
- Klimesch, W., Sauseng, P., Hanslmayr, S., 2007. EEG alpha oscillations: the inhibition timing hypothesis. *Brain Research Reviews* 53, 63–88.
- Kok, A., 1986. Effects of degradation of visual stimulation on components of the event-related potential (ERP) in go/nogo reaction tasks. *Biological Psychology* 23, 21–38.
- Kraut, M.A., Kremen, S., Segal, J.B., Calhoun, V., Moo, L.R., Hart, J., 2002. Object activation from features in the semantic system. *Journal of Cognitive Neuroscience* 14, 24–36.
- Luu, P., Tucker, D.M., Makeig, S., 2004. Frontal midline theta and the error-related negativity: neurophysiological mechanisms of action regulation. *Clinical Neurophysiology* 115, 1821–1835.
- Maguire, M.J., Brier, M.R., Moore, P.S., Ferree, T.C., Ray, D., Mostofsky, S., Hart, J., Kraut, M.A., 2009. Influence of perceptual and semantic categorization on inhibitory processing as measured by the N2-P3 response. *Brain and Cognition* 71, 196–203.
- Maguire, M.J., Brier, M.R., Ferree, T.C., 2010. Differences in EEG theta and alpha responses reveal qualitative differences in processing taxonomic and thematic semantic relationships. *Brain and Language* 114, 16–25.
- Mandler, J.M., 2000. Perceptual and conceptual processes in infancy. *Journal of Cognition and Development* 1, 3–36.
- Mostofsky, S.H., Schafer, J.G.B., Abrams, M.T., Goldberg, M.C., Flower, A.A., Boyce, A., Courtney, S.M., Calhoun, V.D., Kraut, M.A., Denckla, M.B., Pekar, J.K., 2003. fMRI evidence that the neural basis of response inhibition is task-dependent. *Cognitive Brain Research* 17, 419–430.
- Nieuwenhuis, S., Yeung, N., Wildenberg, W.V.D., Ridderinkhof, R., 2003. Electrophysiological correlates of anterior cingulate function in a go/no-go task: effects of response conflict and trial type frequency. *Cognitive, Affective, and Behavioral Neurology* 3, 17–26.
- Nieuwenhuis, S., Yeung, N., Cohen, J.D., 2004. Stimulus modality, perceptual overlap, and the go/no-go N2. *Psychophysiology* 41, 157–160.
- Nunez, P., 1981. *Electrical Fields of the Brain*. Oxford University Press.
- Nunez, P.L., Srinivasan, R., Westdrop, A.F., Wijesinghe, R.S., Tucker, D.M., Silberstein, R.B., Cadusch, P.J., 1997. EEG coherence I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalography and Clinical Neurophysiology* 103, 499–515.
- Perrin, F., Pernier, J., Bertrand, O., Echallier, J.F., 1989. Spherical splines for scalp potential and current density mapping. *Electrophysiology and Clinical Neurophysiology* 72, 184–187.
- Pfurtscheller, G., Stancak, A., Neuper, C., 1996. Event-related synchronization (ERS) in the alpha band – an electrophysiological correlate of cortical idling: a review. *International Journal of Psychophysiology* 24, 39–46.
- Picton, T.W., Stuss, D.T., Alexander, M.P., Shallice, T., Binns, M.A., Gillingham, S., 2007. Effects of focal frontal lesions on response inhibition. *Cerebral Cortex* 17, 826–838.
- Raghavachari, S., Lisman, J.E., Tully, M., Madsen, J.R., Bromfield, E.B., Kahana, M.J., 2006. Theta oscillations in human cortex during a working-memory task: evidence for local generators. *Journal of Neurophysiology* 95, 1630–1638.
- Rappelsberger, P., Petsche, H., 1988. Probability mapping: power and coherence analyses of cognitive processes. *Brain Topography* 1, 46–54.
- Ridderinkhof, K.R., van den Wildenberg, W.P.M., Segalowitz, S.J., Carter, C.S., 2004. Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition* 56, 129–140.
- Roche, R.A.P., Dockree, P.M., Garavan, H., Foxe, J.J., Robertson, I.H., O'Mara, S.M., 2004. EEG alpha power changes reflect response inhibition deficits after traumatic brain injury (TBI) in humans. *Neuroscience Letters* 362, 1–5.
- Rubia, K., Russell, T., Overmeyer, S., Brammer, M.J., Bullmore, E.T., Sharma, T., Simmons, A., Williams, S.C.R., Giampietro, V., Andrew, C.M., Taylor, E., 2001. Mapping motor inhibition: conjunctive brain activations across different versions of Go/No-Go and stop tasks. *Neuroimage* 13, 250–261.
- Sauseng, P., Klimesch, W., Schabus, M., Doppelmayr, M., 2005. Fronto-central EEG coherence in theta and upper alpha reflect central executive functions in working memory. *International Journal of Psychophysiology* 57, 97–103.
- Sauseng, P., Klimesch, W., Gruber, W.R., Hanslmayr, S., Freunberger, R., Doppelmayr, M., 2007. Are event-related potential components generated by phase resetting of brain oscillations? A critical discussion. *Neuroscience* 146, 1435–1444.
- Sawaki, R., Katayama, J., 2009. Difficulty of discrimination modulates attentional capture by regulating attentional focus. *Journal of Cognitive Neuroscience* 29, 359–371.
- Schmiedt, C., Meistrowitz, A., Schendemann, G., Herrmann, M., Basar-Eroglu, C., 2005. Theta and alpha oscillations reflect differences in memory strategy and visual discrimination performance in patients with Parkinson's disease. *Neuroscience Letters* 388, 138–143.
- Simmonds, D.J., Pekar, J.J., Mostofsky, S.H., 2007. Meta-analysis of Go/No-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia* 46, 224–232.
- Srinivasan, R., Nunez, P.L., Silberstein, R.B., 1998. Spatial filtering and neocortical dynamics: estimates of EEG coherence. *IEEE Transactions on Biomedical Engineering* 45, 814–826.
- van Winsun, W., Sergeant, J., Geuze, R., 1984. The functional significance of event-related desynchronization of alpha rhythm in attentional and activating tasks. *Electroencephalography and Clinical Neurophysiology* 58, 519–524.
- von Stein, A., Chiang, C., König, P., 2000. Top-down processing mediated by interareal synchronization. *Proceedings of the National Academy of Science* 97, 14748–14753.
- Weiss, S., Mueller, H.M., 2003. The contribution of EEG coherence to the investigation of language. *Brain and Language* 85, 325–343.
- Yamanaka, K., Yamamoto, Y., 2010. Single-trial EEG power and phase dynamics associated with voluntary response inhibition. *Journal of Cognitive Neuroscience* 22, 714–727.